



Rosanna Huotari

## **Puuston tilavuuden arvioiminen puulajeittain käyttäen avointa satelliittiaineistoa**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi  
diplomi-insinöörin tutkintoa varten.

Espoossa 29.4.2018

Valvoja: Prof. Miina Rautiainen

Ohjaaja: TkT Lauri Häme

---

**Tekijä** Rosanna Huotari

---

**Työn nimi** Puuston tilavuuden arvioiminen puulajeittain käyttäen avointa satelliittiaineistoa

---

**Koulutusohjelma** Master's Programme in Geoinformatics

---

**Työn valvoja** Professori Miina Rautiainen

---

**Työn ohjaaja(t)** Tekniikan tohtori Lauri Häme

---

**Päivämäärä** 29.4.2018

**Sivumäärä** 57 + 13

**Kieli** Suomi

---

## Tiivistelmä

Tämä tutkimus perustui avoimeen satelliittiaineistoon ja tarkoituksena oli selvittää Sentinel-2A-satelliitin kanavien soveltuvuutta puuston tilavuuden arvioinnissa pohjoiseen havumetsävyöhykkeeseen kuuluvalla Etelä-Suomen alueella. Tutkimus jakautui kirjallisuuskatsaukseen sekä kokeelliseen tutkimukseen. Kirjallisuuskatsauksessa taustoitettiin metsän heijastusspektriä, Sentinel-2-satelliitin operaatiota, ominaisuuksia ja aineiston esikäsittelyä sekä metsän mittausta maastossa ja satelliittikuvaan pohjautuvilla menetelmillä. Kokeellisen tutkimuksen referenssiaineistona käytettiin Etelä-Suomen alueelta kerättyjä koealoja, joita oli yhteensä 821 kappaletta. Koko aineisto jaettiin koealalla kasvavan valtapuulajin tilavuuden prosenttiosuuksien mukaan seuraaviin luokkiin: mänty, kuusi, lehtipuu, havusekametsä ja sekametsä. Tilavuutta arvioitiin lineaarisen regressioanalyysin avulla puulajiluokittain ja selitettävänä muuttujana käytettiin sekä tilavuutta että tilavuuden luonnollista logaritmia.

Tilavuuden luonnollisen logaritmin avulla saavutettiin parempia tuloksia, sillä sen suhde Sentinel-2-kanavien kanssa oli selkeämmin lineaarinen. Tulokset osoittivat Sentinel-2-satelliitin punaisen reunan kanavien soveltuvan parhaiten puuston tilavuuden arvioimiseen. Koko aineistolle parhaimman tuloksen tuotti punaisen reunan ensimmäiseen (B5) ja toiseen (B6) kanavaan perustuva regressiomalli (nRMSE 63,0 %). Paras tilavuusmalli männylle, kuuselle, lehtipuulle ja sekametsälle saavutettiin punaisen reunan ensimmäisellä kanavalla (B5), ja vastaavassa järjestyksessä puulajien nRMSE-arvot olivat 49,8 %, 47,2 %, 103,2 % ja 77,6 %. Havusekametsän paras regressiomalli syntyi vihreän valon aallonpituusalueen kanavalla (B3) sekä punaisen reunan toisella kanavalla (B6) ja sen nRMSE oli 49,0 %.

Havupuuvaltaisten metsien tilavuuden arvioiminen oli tarkempaa kuin lehtipuuvaltaisten metsien. Tässä tutkimuksessa Sentinel-2-satelliittiaineistoon perustuvan tilavuusarvioinnin tarkkuus oli samaa luokkaa aikaisemmin tutkittujen satelliittiaineistojen kanssa. Tässä tutkimuksessa käytettyjen satelliittikuvien alueelle laskettiin tilavuuskartta, joka vastasi pääosin monilähteisen valtakunnan metsien inventoinnin tuottamia tilavuusarvioita.

---

**Avainsanat** Sentinel-2, tilavuus, regressioanalyysi, metsä, havumetsävyöhyke

---



---

**Author** Rosanna Huotari

---

**Title of thesis** Estimating growing stock volume of different tree species based on open satellite data

---

**Degree programme** Master's Programme in Geoinformatics

---

**Thesis supervisor** Professor Miina Rautiainen

---

**Thesis advisor(s)** Doctor of Science (Tech.) Lauri Häme

---

**Date** 29.4.2018

**Number of pages** 57 + 13

**Language** Finnish

---

## **Abstract**

This master thesis was based on open satellite data and its aim was to find out which Sentinel-2A-satellite bands are suitable to estimate forest volume. The study area was in Southern Finland which belongs to boreal forest zone. The study was divided into literature review and empirical research. The literature review discussed forest reflectance spectrum, Sentinel-2-satellite operation, characteristics and its data preprocessing, forest mensuration in the field and based on satellite data. In the empirical research the used reference data was collected from Southern Finland. There were 821 sample plots altogether. The reference data was divided into following classes: pine, spruce, deciduous tree, mixed coniferous forest and mixed forest, based on percentage of growing stock volume of major tree in the sample plot. Linear regression analysis was used to estimate growing stock volume in each class. The dependent variable was either growing stock volume or natural logarithm of growing stock volume.

Natural logarithm of growing stock volume as dependent variable performed better, because of stronger linear relationship between natural logarithm of growing stock volume and Sentinel-2-bands. The results indicated that red edge bands estimated better growing stock volume. The best results for the whole reference data were achieved using both the first (B5) and the second (B6) red edge bands (nRMSE 63,0 %). The best volume estimation in classes of pine, spruce, deciduous tree and mixed forest was achieved with the first red edge band (B5) and their nRMSE results were 49,8 %, 47,2 %, 103,2 % ja 77,6 %, respectively. The best volume estimation for mixed coniferous forest was achieved with band of green light (B3) and the second red edge band (B6) and its nRMSE was 49,0 %.

Estimation of growing stock volume of forest consisting mainly coniferous trees was more accurate than forest consisting mainly deciduous trees. In this study growing stock volume estimation based on Sentinel-2-satellite data was as accurate as volume estimations obtained in earlier studies. In this study a wall-to-wall map of growing stock volume was calculated for the area of used satellite images. The map was mainly similar with the volume estimation based on multi-source national forest inventory.

---

**Keywords** Sentinel-2, volume, regression analysis, forest, boreal forest zone

---

## Alkusanat

*Tämän tutkimuksen aihe tarjottiin Satellio Oy:n toimesta. Satellio Oy halusi selvittää, kuinka hyvin Sentinel-2-satelliitin aineisto soveltuu puuston tilavuuden arvioimiseen. Tutkimustyöni tarkempi raja-  
aus muodostui yhdessä ohjaajani TkT Lauri Hämeen kanssa käytyjen keskusteluiden pohjalta. Haluan kiittää Satellio Oy:tä työni mahdollistamisesta sekä ohjaajaani Lauri Hämettä ohjauksesta ja korjausehdotuksista työni aikana. Lisäksi haluan kiittää valvojaani professori Miina Rautiaista hyvistä neuvoista ja palautteesta työni kehittämiseksi. Iso kiitos kuuluu myös perheelleni ja Joni Salolle tuesta ja kannustuksesta opintojeni aikana sekä diplomityön loppuun saattamisessa.*

Espoo 29.4.2018

Rosanna Huotari

# Sisällysluettelo

Tiivistelmä	
Abstract	
Alkusanat	
Sisällysluettelo .....	5
Merkinnät .....	6
Lyhenteet .....	7
1 Johdanto .....	8
1.1 Tutkimuksen tausta .....	8
1.2 Tutkimuksen tavoitteet ja tutkimusmenetelmät .....	9
1.3 Tutkimuksen rajausta ja rakenne .....	9
2 Kirjallisuuskatsaus .....	10
2.1 Metsän heijastusspektri .....	10
2.2 Sentinel-2 .....	13
2.2.1 Operaatio ja ominaisuudet .....	13
2.2.2 Sentinel-2-kuvan esiprosessointi .....	14
2.3 Metsän mittaus maastossa .....	16
2.4 Puuston tilavuuden arvioiminen satelliittikuvalta .....	19
2.4.1 Tilavuuden arviointimenetelmiä .....	19
2.4.2 Regressioanalyysi .....	23
3 Aineisto ja menetelmät .....	26
3.1 Maastokoealat .....	26
3.2 Sentinel-2-satelliittikuva .....	27
3.3 Regressiomalli puuston tilavuuden ennustamiseksi .....	29
3.4 Tilavuuskartta ja MVMI-tuote .....	31
4 Tulokset ja niiden tarkastelu .....	32
4.1 Pearsonin korrelaatiokertoimet .....	32
4.2 Regressioanalyysin tulokset .....	36
4.2.1 Selitettävänä muuttujana tilavuus .....	36
4.2.2 Selitettävänä muuttujana tilavuuden luonnollinen logaritmi .....	38
4.3 Regressioanalyysin tulosten tarkastelu ja mallin valinta .....	40
4.4 Validoinnin tulokset ja niiden tarkastelu .....	43
4.5 Tilavuuskartan vertailu MVMI-tuotteeseen .....	44
5 Yhteenveto .....	50
Lähdeluettelo .....	51
Liiteluettelo .....	57
Liitteet	

## Merkinnät

D	[cm]	metsikön keskiläpimitta
E	[W/sr]	irradianssi
G	[m <sup>2</sup> /ha]	metsikön pohjapinta-ala
H	[m]	metsikön keskipituus
L	[W/m <sup>2</sup> /sr]	radianssi
M	[W/m <sup>2</sup> ]	kohteen pinnasta lähtevä sähkömagneettinen säteily
N	[kpl/ha]	metsikön runkoluku
T	[a]	metsikön puuston ikä
V	[m <sup>3</sup> /ha]	metsikön puuston tilavuus
d <sub>1,3</sub>	[cm]	puun rinnankorkeusläpimitta
f	[-]	puun rinnankorkeusmuotoluku
g	[m <sup>2</sup> ]	puun pohjapinta-ala
h	[m]	puun pituus
v	[m <sup>3</sup> ]	puun tilavuus

## Lyhenteet

ANOVA	Analysis of Variance, varianssianalyysi
AOT	Aerosol Optical Thickness, aerosolien optinen paksuus
BOA	Bottom of Atmosphere, ilmakehän alapuoleinen
DEM	Digital Elevation Model, digitaalinen korkeusmalli
DN	Digital Number, digitaalinen kokonaisluku
DTED	Digital Terrain Elevation Data
DVI	Difference Vegetation Index
ESA	The European Space Agency, Euroopan avaruusjärjestö
ETM+	Enhanced Thematic Mapper Plus
EVI	Enhanced Vegetation Index
GMES	The Global Monitoring for Environment and Security
IPVI	Infrared Percentage Vegetation Index
Luke	Luonnonvarakeskus
LUT	Look-Up Table, hakutaulukko
MdAE	Median Absolute Error, absoluuttinen mediaanivirhe
MSI	Multi-Spectral Imager, multispektraalinen kuvaussensori
MVMI	Monilähteinen valtakunnan metsien inventointi
NIR	Near-infrared, lähi-infrapuna
PNS-menetelmä	Pienimmän neliösumman menetelmä
RMSE	Root Mean Square Error, keskineliövirheen neliöjuuri
SAVI	Soil-Adjusted Vegetation Index
SC	Scene Classification, näkymän luokittelu
SR	Simple Ratio
SSE	Sum of squared error, selittämättömien poikkeamien summa
SSR	Sum of squares regression, mallin selittämien poikkeamien summa
SST	Sum of square total, kokonaispoikkeamien neliöiden summa
SWIR	Shortwave infrared, lyhytaaltainen infrapuna
Syke	Suomen ympäristökeskus
TM	Thematic Mapper
TOA	Top of Atmosphere, ilmakehän yläpuoleinen
UTM	Universal Transverse Mercator, karttaprojektio
VIF	Variance Inflation Factor
VMI	Valtakunnan metsien inventointi
WGS84	World Geodetic System 1984, tasokoordinaattijärjestelmä
WV	Water Vapour, vesihöyry
k-MSN	k-Most Similar Neighbours, k:n samanlaisimman naapurin menetelmä
k-NN	k-Nearest Neighbour, k:n lähimmän naapurin menetelmä
nMdAE	normalized Median Absolute Error, normalisoitu absoluuttinen mediaanivirhe
nRMSE	normalized Root Mean Square Error, normalisoitu keskineliövirheen neliöjuuri

# 1 Johdanto

## 1.1 Tutkimuksen tausta

Suomessa on suotuisat olosuhteet metsänkasvulle, sillä Golf-virta lämmittää ilmastoa ja kesällä valoa on paljon. Suomen metsät kuuluvat lähes kokonaan pohjoiseen havumetsävyöhykkeeseen, jonka tyypillisimmät puulajit ovat kuusi ja mänty. Edellä mainittujen havupuiden lisäksi lehtipuista koivut, lepät, pajut sekä haavat menestyvät hyvin. (Lindberg ja Nygren 2014.)

Puuston vuosittainen kasvu on ollut Suomessa 1970-luvun jälkeen hakuista ja luontaista kuolemista suurempaa (Luonnonvarakeskus 2017). Vuodesta 1920 vuoteen 2016 puuston määrä on kasvanut 1,7-kertaiseksi, mitä suurelta osin selittää tehostunut metsänhoito 1960-1970-luvuilla (Palosuo 1979). Puuston määrän kasvamisen myötä kiireelliset metsänhoitotarpeet ovat myös lisääntyneet (Luonnonvarakeskus 2017).

Metsänomistajilla sekä metsäorganisaatioilla on jatkuva tarve päivitetylle metsätiedolle. Kuitenkin perinteinen maastossa suoritettava metsäinventointi on kallista ja päivitysnopeus liian hidas. (Astola et al. 2004.) Osana ongelman ratkaisua on mahdollista käyttää kaukokartoitussatelliitteja, joiden avulla tietoa voidaan kerätä jatkuvasti. Päivitysnopeuden kasvun lisäksi tiedon kerääminen on kustannustehokasta myös laajoilta alueilta. (Chrysafis et al. 2017, Hovi et al. 2017.)

Metsätaloudessa yksi metsää kuvaavista keskeisimmistä tunnusluvuista on puuston tilavuus ( $\text{m}^3/\text{ha}$ ) (Paananen et al. 2009). Satelliittikuvaan pohjautuvia tilavuusarvioita on tutkittu muun muassa seuraavissa maissa: Suomi (Hyypä et al. 2000, Astola et al. 2004, Mäkelä ja Pekkarinen 2004, Peuhkurinen et al. 2008), Ruotsi (Holmgren et al. 2000), Kanada (Hall et al. 2006), Iran (Mohammadi et al. 2010), Turkki (Günlü et al. 2015), Kreikka (Chrysafis et al. 2017), Kiina (Huiyan et al. 2006) ja Uusi-Seelanti (Trotter et al. 1997). Edellä mainituissa tutkimuksissa on hyödynnetty Landsat-, SPOT- tai IKONOS-kuvaa. Näiden satelliittikuvien spatiaalinen resoluutio vaihtelee sensorin mukaan kymmenistä metreistä alle metriin. Landsat- ja SPOT-satelliittikuvia vertailevan tutkimuksen mukaan SPOT-satelliitin kuvat sopivat puuston tilavuuden arvioimiseen Landsat-satelliitin kuvia paremmin (Hyypä et al. 2000).

SPOT- ja Landsat-missioiden rinnalle on suunniteltu Sentinel-2-satelliittipari, jonka tarkoituksena on täydentää multispektrihavaintoja ja varmistaa havaintojen jatkuvuus (Drusch et al. 2012). Sentinel-2-satelliittiparin ensimmäinen satelliitti Sentinel-2A laukaistiin avaruuteen vuonna 2015 ja toinen, Sentinel-2B, vuonna 2017. Sentinel-2-satelliittiparin kuvaustaajuus on 5 päivää ja niiden kanavat kattavat näkyvän valon, punaisen reunan, lähi-infrapun ja lyhytaaltoisen infrapun aallonpituusalueita. Yhteensä kanavia on 13 ja niiden spatiaalinen resoluutio vaihtelee kanavakohtaisesti 10-60 metrin välillä. (ESA 2017.) Merkittävänä erona edellisessä kappaleessa mainittuihin satelliitteihin on Sentinel-2-satelliittien kolme uutta kanavaa punaisen reunan alueella. Näitä kanavia pidetään lupaavina kasvillisuuden kaukokartoituksessa, sillä niiden odotetaan parantavan useiden biofysikaalisten muuttujien kuten klorofyllipitoisuuden arviointitarkkuutta. Biofysikaaliset muuttujat korreloivat voimakkaasti esimerkiksi puuston tilavuuden kanssa. (Korhonen et al. 2017.) Näin ollen ainakin teoreettisesti Sentinel-2-satelliittikuvahavainnot soveltuvat hyvin metsätalouden sovelluksiin.



Puuston tilavuuden arvioinnissa Sentinel-2-kuvahavaintojen käyttö on vielä vähäistä satelliittiparin nuoren iän vuoksi. Chrysafisin et al. (2017) alustavassa tutkimuksessa arvioitiin Sentinel-2-satelliitin soveltuvuutta Välimeren puuston tilavuuden arvioinnissa. Vertailukohteenä heillä oli Landsat 8-satelliittikuva. Tutkimuksessa havaittiin tuloksien parantuvan, kun käytettiin lähi-infrapunan sijaan punaisen reunan kanavia. Kaiken kaikkiaan tarkkuuksissa ei ollut suurta eroa Landsat 8- ja Sentinel-2-kuvien välillä.

## **1.2 Tutkimuksen tavoitteet ja tutkimusmenetelmät**

Tämän tutkimuksen tavoitteena oli selvittää, kuinka hyvin Sentinel-2-satelliittiaineisto soveltuu puuston tilavuuden arviointiin pohjoisella havumetsävyöhykkeellä. Tutkimus koostui sekä kirjallisuuskatsauksesta että kokeellisesta tutkimuksesta. Kirjallisuuskatsauksessa esiteltiin metsän heijastusspektriin vaikuttavia tekijöitä, Sentinel-2-satelliittien operaatio, ominaisuuksia ja aineiston esikäsittely sekä maastomittauksiin ja satelliittikuvaan pohjautuvia puuston arviointimenetelmiä. Kokeellisessa tutkimuksessa puuston tilavuutta arvioitiin puulajeittain – mänty, kuusi, lehtipuu, havusekametsä ja sekametsä. Tilavuusarvioinnissa käytettiin lineaarista regressioanalyysiä, jossa selitettävänä muuttujana käytettiin sekä puuston tilavuutta että tilavuuden luonnollista logaritmia.

Tämän työn tutkimuskysymys oli:

Kuinka hyvin Sentinel-2-satelliittikuva soveltuu puuston tilavuuden arviointiin?

Tutkimuskysymystä tarkentavia alakysymyksiä olivat:

- Mitkä Sentinel-2-kanavat soveltuvat puuston tilavuuden arviointiin?
- Mikä on regressioanalyysin tarkkuus puuston tilavuuden arvioinnissa?
- Mikä on tilavuusarvioinnin tarkkuus puulajeittain?
- Kuinka tarkasti tämän tutkimuksen tilavuusarvio vastaa monilähteisen valtakunnan metsien inventoinnin tilavuusarviota?

## **1.3 Tutkimuksen raja- ja rakenne**

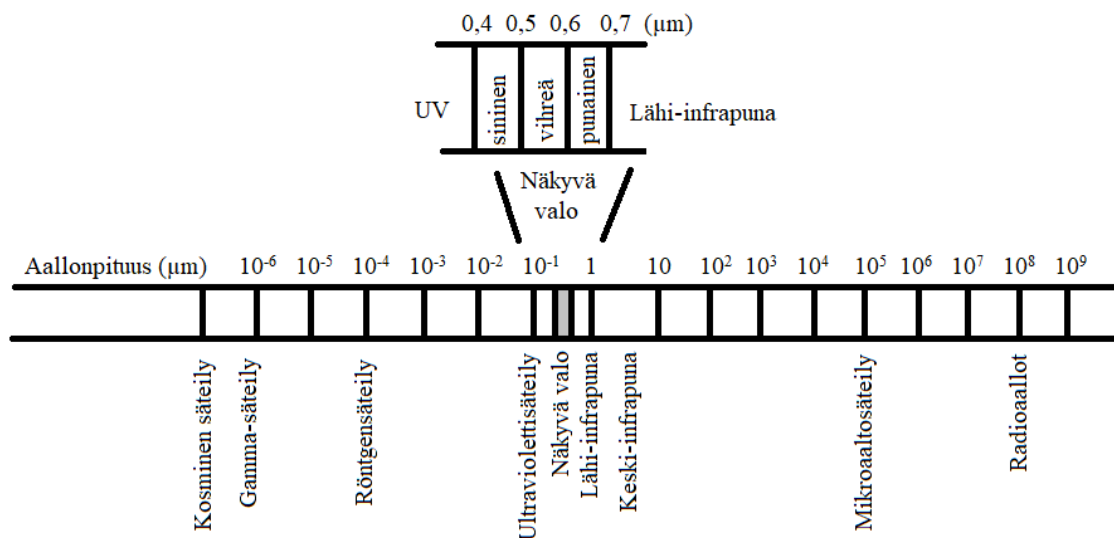
Tutkimustyö rakentuu viidestä luvusta, joista ensimmäinen on johdanto. Toinen luku on kirjallisuuskatsaus, jossa on neljä alalukua. Luvussa 2.1 käsitellään metsän heijastusspektriä. Luvussa 2.2 esitellään Sentinel-2-satelliittien operaatio, ominaisuuksia ja kuva-aineiston esikäsittely. Tutkimuksen kokeellisessa osiossa käytetään ainoastaan Sentinel-2-satelliittikuvaa, mutta tuloksia vertaillaan myös muita satelliittikuvia käyttäviin tutkimuksiin. Luvussa 2.3 esitellään lyhyesti puuston maastomittaus, jonka jälkeen luvussa 2.4 esitellään menetelmiä puuston tilavuuden arvioimiseksi satelliittikuvalta. Satelliittikuvaan pohjautuvissa menetelmissä tarkastellaan sekä pohjoisella havumetsävyöhykkeellä että muilla kasvillisuusvyöhykkeellä toteutettuja tutkimuksia. Menetelmistä esitellään tarkemmin regressioanalyysi, joka on tämän tutkimuksen kokeellisen osion menetelmä.

Kokeellinen tutkimus käsitellään luvuissa 3 ja 4. Luvussa 3 käydään läpi puuston tilavuusmallin kehittämisessä käytetyt aineistot ja menetelmät. Puuston tilavuusmallin referenssiaineisto ja satelliittikuvat ovat Etelä-Suomen alueelta. Tämän jälkeen luvussa 4 esitellään tutkimustulokset sekä niiden tarkastelu. Lopuksi luvussa 5 on yhteenveto tämän tutkimuksen tuloksista.

## 2 Kirjallisuuskatsaus

### 2.1 Metsän heijastusspektri

Passiivisen, optisen kaukokartoitussatelliitin toiminta perustuu auringon emittoimaan sähkömagneettiseen säteilyyn, joka on heijastunut maanpinnalla sijaitsevasta kohteesta (Wulder 1998). Instrumentin sensori mittaa kohteen heijastamaa säteilytehoa eli radianssia ( $L$  [ $W/m^2/sr$ ]) sähkömagneettisen spektrin eri aallonpituusalueilla (Kuva 1). Yleensä nämä alueet sijaitsevat ihmissilmän havaitseman näkyvän valon sekä lähi-infrapun (engl. near-infrared, NIR) ja lyhyt aaltoisen infrapun (engl. short-wave infrared, SWIR) aallonpituusalueilla (Asner 1998).

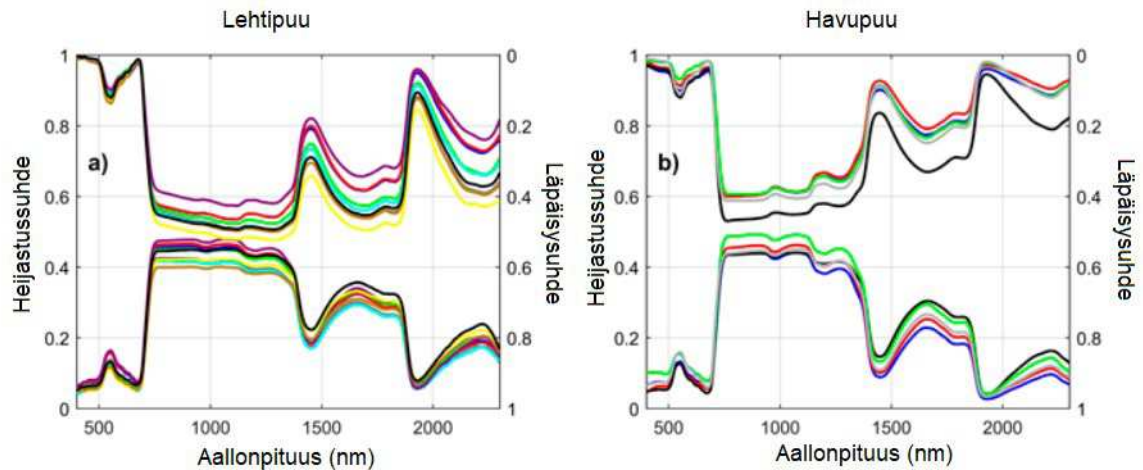


Kuva 1. Sähkömagneettinen spektri. (mukaillen Lillesand et al. 2008)

Heijastussuhde (engl. reflectance) on suhde kohteen pinnasta lähtevän ( $M$  [ $W/m^2$ ], engl. radiant exitance) ja siihen saapuvan säteilyn irradianssin ( $E$  [ $W/m^2$ ], engl. irradiance) välillä. Heijastussuhteen määrään vaikuttavat kohteen heijastaman säteilyn lisäksi ilmakehän olosuhteet, kohteen ympäristö sekä topografia, joiden vaikutus on erilainen eri aallonpituusalueilla. Heijastussuhde vaihtelee samalla kohteella myös valaistuksen ja sensorin mittausgeometrian mukaan. Pinnan heijastusominaisuuksia mitattaessa käytetään käsitettä heijastussuhdetekijä (engl. reflectance factor). Heijastussuhdetekijä on suhde tutkitavan kohteen pinnasta heijastuneen säteilyvuon sekä ideaalin diffuusin pinnan heijastaman säteilyvuon välillä, kun molemmissa tapauksissa olosuhteet ovat samanlaiset. (Schaeppman-Strub et al. 2006.)

Sähkömagneettisen säteilyn vuorovaikutusta kohteen kuten metsän kanssa kuvataan heijastusspektrillä, jossa heijastussuhde esitetään aallonpituuden funktiona. Yleisesti metsän heijastusspektri vastaa kasvillisuuden heijastusspektrin piirteitä, jolle tyypillistä on vihreän valon alueella esiintyvä heijastuspiikki sekä punainen reuna. Vihreän valon heijastuspiikki on havaittavissa noin 550 nm kohdalla (Kuva 2a ja b). Punaisen ja sinisen valon alueilla klorofyllit absorboivat säteilyä, jolloin takaisin heijastunut säteily on vihreää aallonpituusalueelta huomattavasti alhaisempi. Punainen reuna esiintyy punaisen ja lähi-infrapun aallonpituusalueiden välillä noin 700 nm kohdalla (Kuva 2a ja b). Kasvillisuus heijastaa voimak-

kaasti lähi-infrapunaa aallonpituuksia, johon vaikuttaa muun muassa lehtien sisäinen rakenne ja niiden vähäinen absorptio. Aallonpituuden ollessa 1400-1900 nm välillä heijastussuhde on alhainen, joka aiheutuu lehden sisältämän veden voimakkaasta absorptiosta. (Knippling 1970.)



Kuva 2. (a) Lehti- ja (b) havupuiden heijastus- ja läpäisysuhde. Alempi käyrä kuvaa heijastussuhdetta ja ylempi läpäisysuhdetta. (mukaillen Hovi et al. 2017)

Metsän heijastusspektriin vaikuttavista tekijöistä tarvitaan tietoa, jotta kaukokartoitussatelliitin instrumentin havaintoja voidaan tulkita. Sensorin mittaamaan radianssiin vaikuttaa puun osien biokemiallisten ja metsän rakenteellisten ominaisuuksien (Asner 1998) lisäksi vuodenaika (Hovi et al. 2017), ilmakedä (Wulder 1998), spatiaalinen resoluutio, valaistus- ja kuvauskulma (Rautiainen et al. 2018).

Biokemialliset ominaisuudet kuten esimerkiksi klorofyllit ja karotenoidit määrittävät yksittäisten lehtien, neulasten, runkojen ja oksien optisia ominaisuuksia (Lukeš et al. 2013). Lukeš et al. (2013) havaitsivat tutkimuksessaan, että männyn ja kuusen neulasten läpäisysuhde oli selkeästi alhaisempi kuin heijastussuhde. Koivulla vuorostaan läpäisy- ja heijastussuhde vastasivat toisiaan. Hovi et al. (2017) tutkivat boreaalisen vyöhykkeen 25 puulajin neulasten ja lehtien spektrisiä ominaisuuksia. Kuvassa 2 havainnollistetaan tutkimuksessa havaittuja lehti- ja havupuiden heijastus- sekä läpäisysuhteen eroa. Tutkimuksessa tarkasteltiin useita puulajeja, jotka näkyvät kuvassa 2 erivärisinä käyriä. Havu- ja lehtipuiden lajien sisäiset erot kasvoivat, kun lähestyttiin SWIR-aallonpituusalueita (noin 1300-2500 nm). Lisäksi punaisen reunan alueella havupuiden erot olivat melko hyvin havaittavissa. Erot eivät olleet yhtä selkeitä näkyvän valon ja lähi-infrapunaa aallonpituusalueilla. Esimerkiksi Sentinel-2-satelliitissa kanavia on sekä punaisen reunan että SWIR-aallonpituusalueella, jolloin lajien erottaminen toisistaan helpottuu. Hovi et al. (2017) havaitsivat myös, että SWIR- ja punaisen reunan aallonpituusalueet korreloivat heikosti muiden aallonpituusalueiden kanssa, mikä lisää näiden aallonpituusalueiden tärkeyttä.

Kaukokartoitetun aineiston heijastussuhteeseen vaikuttavat neulasten ja lehtien lisäksi myös puun muut elementit kuten oksat ja runko. Tämän seurauksena kaukokartoitusaineistossa metsän heijastussuhde on pienempi kuin pelkästään lehdistä tai neulasista mitattuna. Kuitenkin lehtien ja neulasten spektrien ominaisuuksia on tutkittu enemmän kuin muita puun osia kuten kuorta. (Rautiainen et al. 2018.) Puunkuoren heijastusspektri vaihtelee lajikohtaisesti,

mutta yleisesti kuoren heijastussuhde kasvaa monotoonisesti näkyvän valon ja NIR-aallonpituusalueiden läpi (Hall et al. 1992).

Yksittäisten puun elementtien kuten neulasten, lehtien, rungon ja oksien lisäksi kaukokartoitushavaintoihin vaikuttavat metsikön rakenteelliset ominaisuudet kuten tiheys, latvuspeitto sekä aluskasvillisuus. Aluskasvillisuudella on merkittävä vaikutus koko metsän heijastussuhteeseen. Latvuspeitoltaan harvoissa metsissä satelliitin sensori havaitsee enemmän aluskasvillisuutta kuin tiheään latvuspeiton metsissä (Eriksson et al. 2006). Suurin vaikutus aluskasvillisuudella on näkyvän valon aallonpituusalueella, jossa kasvukauden aikana on havaittavissa suuria muutoksia. Punaisen reunan ja NIR-aallonpituusalueilla aluskasvillisuuden vaikutus on vähäisempi koko kasvukauden ajan. (Rautiainen ja Lukeš 2015.)

Vuodenaika vaikuttaa aluskasvillisuuden lisäksi myös lehti- (Möttus et al. 2014) ja havupuiden (Rautiainen et al. 2018) heijastusspektriin. Havumetsävyöhykkeellä toukokuussa lehtipuiden näkyvän valon vihreän, punaisen sekä punaisen reunan aallonpituusalueiden absorptio kasvaa, kun taas lokakuussa on havaittavissa vähenemistä. NIR- sekä SWIR-aallonpituusalueilla heijastus kasvaa keväällä nopeasti ja jatkuu hitaampana kasvuna läpi kesän ja syksyn. Havupuilla kesä- ja syyskuun välillä on havaittavissa muutoksia NIR-aallonpituusalueen heijastuksessa sen mukaan, onko kyseessä uuden kasvukauden vai aikaisempien vuosien vanhemmat neulaset. Uusien neulasten heijastavuus NIR-aallonpituusalueella kasvaa, kun taas vanhemmilla neulasilla vähenee. (Hovi et al. 2017.)

Sensorin ja kohteen väliin jäävä ilmakehä vaikuttaa merkittävästi havaittuun aineistoon. Ilmakehän pienhiukkaset sekä kaasut muuttavat sähkömagneettista säteilyä sirottamalla tai absorboimalla sitä (Fraser ja Kaufman 1985). Esimerkiksi 1400 ja 1900 nm aallonpituuksien välillä ilmakehän vesihöyry estää maanpinnan kohteen mittaamisen (Asner 1998). Lisäksi sensorin havaitsema säteily ei kuvaa ainoastaan kohteen heijastamaa sähkömagneettista säteilyä, vaan siihen on sekoittunut viereisistä kohteista heijastunutta säteilyä (Korhonen et al. 2017).

Satelliitin sensorin spatiaalinen resoluutio vaikuttaa havaitun heijastussuhteen voimakkuuteen. Satelliitista kaukokartoitetun aineiston spatiaalinen resoluutio vaihtelee sensorin mukaan karkeasta (esim. MODIS, spatiaalinen resoluutio 500 m) korkeaan (esim. Sentinel-2, spatiaalinen resoluutio 20 m). Karkean resoluution satelliitin vastaanottama signaali kuvaa laajempia kokonaisuuksia kuten ekosysteemiä. Korkean resoluution satelliitin signaali vuorostaan kuvaa pienempiä kokonaisuuksia kuten metsikköä. (Rautiainen et al. 2018.) Rautiainen et al. (2018) tutkimuksen mukaan MODIS-satelliittiaineiston heijastussuhde oli korkeampi kuin Sentinel-2-satelliitista havaitulla aineistolla, erityisesti huomattava ero oli infrapuna-aallonpituusalueella. Heijastussuhteiden eroa voi selittää MODIS-satelliitin karkea resoluutio, jolloin yksi pikseli kuvaa useampia maanpinnan tyyppejä.

Valaistus- ja havaintogeometria vaikuttavat metsän osien kuten latvuspeiton valoisten ja varjoisten osien suhteeseen, mikä vuorostaan vaikuttaa sensorin havaitsemaan heijastussuhteeseen (Li ja Strahler 1992). Esimerkiksi havumetsien heijastussuhde vaihtelee katselugeometrian funktiona (Rautiainen et al. 2018). Rautiainen et al. (2018) havaitsivat tutkimuksessaan, että taaksepäin suuntautuvalla kuvausgeometrialla sekä 40-vuotiaan että 80-vuotiaan kuusimetsän heijastussuhde oli suurempi kuin eteenpäin suuntautuvalla kuvausgeometrialla. Vanhemmalla kuusimetsällä kuvausgeometrian muutokset vaikuttivat voimakkaammin. Tätä voi selittää vanhemman kuusimetsän harvempi latvuspeitto ja pidemmät puut, jotka luovat

enemmän vaihtelua latvuston valoisiin ja varjoisiin osiin. Viisto kuvauskulma vähentää myös aluskasvillisuuden vaikutusta kaukokartoitushavaintoihin verrattuna nadiirikuvaukseen (Rautiainen et al. 2018).

## 2.2 Sentinel-2

### 2.2.1 Operaatio ja ominaisuudet

Copernicus (aik. GMES-ohjelma) on Euroopan komission kaukokartoitusohjelma, joka tarjoaa tarkkaa, oikea-aikaista sekä helposti saatavaa informaatiota. Ohjelman tarkoituksena on parantaa ympäristön hallintaa, lisätä ymmärrystä ja sitä kautta lievittää ilmastonmuutoksen vaikutuksia sekä varmistaa siviilien turvallisuutta. Euroopan avaruusjärjestö (engl. European Space Agency, ESA) kehittää Sentinel-satelliittiperhettä, jonka tarkoituksena on vastata Copernicus-ohjelman operatiivisiin tarpeisiin. (ESA 2017.)

Sentinel-perheeseen kuuluu seitsemän missiota, joista tässä tutkimuksessa lähemmin tarkastellaan Sentinel-2-missiota. Sentinel-2 on suunniteltu tarjoamaan korkean resoluution optisia kuvia maatalous-, metsä-, sekä muista maankäyttöalueista. Maanpinnan lisäksi Sentinel-2-missiossa valvotaan rannikoita ja sisävesiä. Missioon kuuluu kaksi satelliittia, joita ensimmäinen Sentinel-2A (Kuva 3) laukaistiin avaruuteen 23.6.2015 ja toinen Sentinel-2B 7.3.2017. (ESA 2017.)



Kuva 3. Sentinel-2-satelliitti kiertoradallaan. (ESA 2017)

Sentinel-2-sarjan satelliitit kulkevat aurinkosynkronisella radalla 786 km korkeudessa 180° etäällä toisistaan. Yhteen kierrokseen menee 100 minuuttia ja näin ollen päivän aikana kertyy 14,3 kierrosta maapallon ympäri. Satelliitin instrumentti kuvaa maapintaa 290 km leveydeltä. Kahdella satelliitilla saadaan katettua koko maapallon pinta 5:ssä päivässä. (ESA 2017.)

Sentinel-2A ja -2B kantavat mukanaan multispektraalista kuvaussensoria (engl. Multi-Spectral Imager, MSI). MSI-sensori kuvaa 13:ta eri kanavaa. Taulukko 1 esittää kanavien aallonpituusalueet ja niiden spatiaalisen erotuskyvyn. Näkyvän valon (B2, B3, B4) ja lähi-infrapun (B8) kanavilla spatiaalinen erotuskyky on 10 m, punaisen reunan (B5, B6, B7), kapean lähi-infrapun (B8a) ja SWIR-aallonpituusalueiden (B11, B12) kanavilla 20 m sekä ilmakehää kuvaavilla kanavilla (B1, B9 ja B10) 60 m. (ESA Sentinel Online 2017.)

*Taulukko 1. Sentinel-2A ja 2B-satelliittien kanavat. (ESA Sentinel Online 2017)*

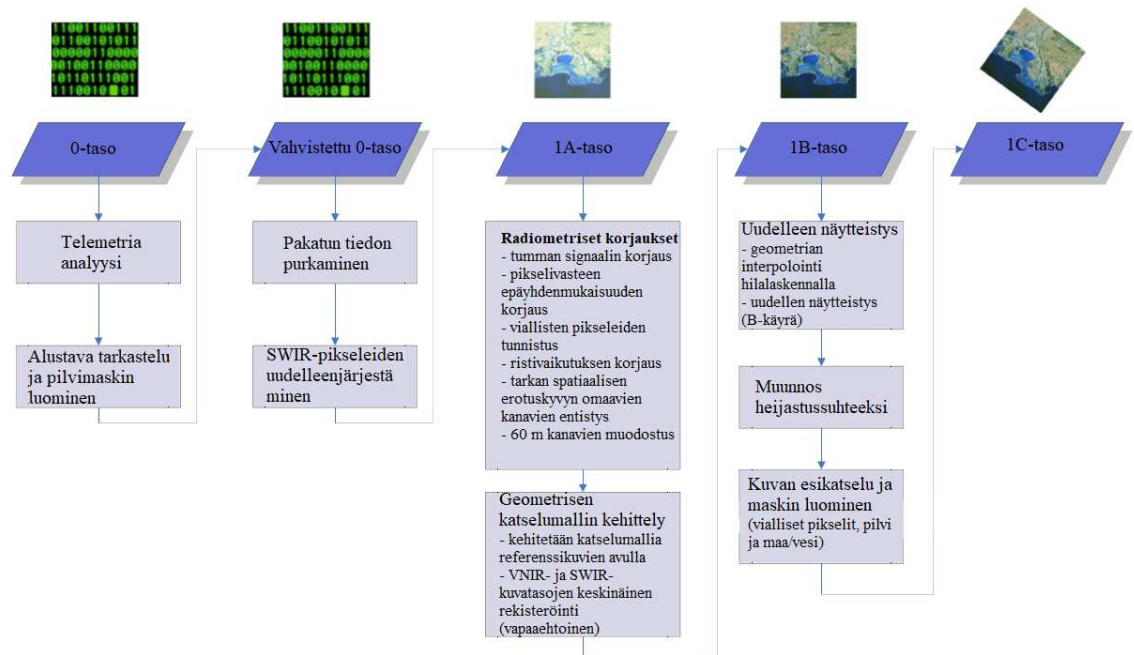
Kanavan numero	Spatiaalinen erotuskyky (m)	S2A		S2B	
		Kanavan lonpituus (nm)	keskiaal- leveys (nm)	Kanavan lonpituus (nm)	keskiaal- leveys (nm)
1	60	443,9	27	442,3	45
2	10	496,6	98	492,1	98
3	10	560,0	45	559	46
4	10	664,5	38	665	39
5	20	703,9	19	703,8	20
6	20	740,2	18	739,1	18
7	20	782,5	28	779,7	28
8	10	835,1	145	833	133
8a	20	864,8	33	864	32
9	60	945,0	26	943,2	27
10	60	1373,5	75	1376,9	76
11	20	1613,7	143	1610,4	141
12	20	2202,4	242	2185,7	238

### 2.2.2 Sentinel-2-kuvan esiprosessointi

Sentinel-2-kuvan esiprosessointi on jaettu eri tasoihin ja sen seurauksena syntyviin tuoteisiin (Kuva 4). Sentinel-2 Ground Segment käsittelee ja prosessoi tasot 0-1C. (ESA Sentinel Online 2017.) Alimman 0-tason tuotteet ovat pakattua raakatietoa, joka sisältää tarvittavan informaation ylemmän tason tuotteiden tuottamiseksi (Gascon et al. 2017).

Alimmalta tasolta seuraavalle 1A-tasolle siirryttäessä suoritetaan nopea prosessointi, jossa kuva-aineisto puretaan, geometrinen malli lasketaan sekä SWIR-pikselit järjestetään uudelleen. 1A-tason pakkaamatonta raakatietoa on mahdollista esikatsella täydessä resoluutiossa kuvankäsittelyohjelmistoilla. (Gatti ja Naud 2017.)

Seuraava 1B-tason tieto muodostuu, kun 1A-tason tuotteelle suoritetaan radiometrinen korjaus. Radiometrisessä korjauksessa sensorin tallentamat digitaaliset kokonaisluvut (engl. digital number, DN) muutetaan radianssiksi. Instrumentin kalibrointikertoimia käytetään tarkassa radiometrisessä mallissa, johon kuuluu sekä tumman signaalin poistaminen että pikselivasteen epäyhdenmukaisuuden korjaus. Radiometrisen korjauksen lopuksi luodaan radiometristä laatua kuvaavat maskit saturoiduille sekä virheellisille pikseleille. Lisäksi muodostetaan 60 m kanavat (B1, B9 ja B10) signaali-kohinasuhteen parantamiseksi. Tässä vaiheessa geometristä mallia on myös kehitetty maaston kontrollipisteiden avulla. (Gascon et al. 2017.) Geometrinen malli liittyy jokaisen pikselin katselusuuntaan ja se yhdistää paikan, asennon, päiväyksen sekä muunnosmatriisit satelliitin, instrumentin, polttotason ja ilmaisimien vertauskehyksien (ent. koordinaatisto) välillä. Automaattinen korrelaatioprosessi korjattavan kuvan referenssikanavan ja referenssikuvan välillä määrää maaston kontrollipisteet, mikä mahdollistaa katselumallin kalibroimisen sekä paikan tai asennon muutosten korjaamisen. (Gascon et al. 2014.) Vaikka malli on lisätty 1B-tason tuotteeseen, niin sitä ei ole vielä sovellettu käytännössä (Gatti ja Naud 2017). 1B-tason tuotetta käytetään kalibrointi-, validointi- ja laadunvalvontatarkoituksiin (Gascon et al. 2017).



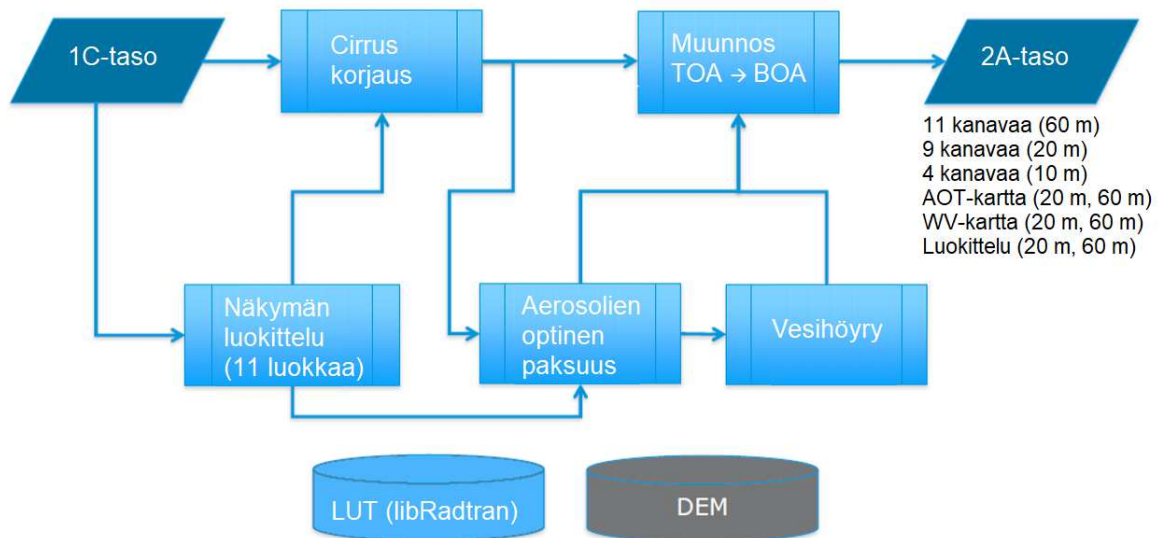
Kuva 4. Sentinel-2-tuotteen prosessointitasot. (mukaillen ESA Sentinel Online 2017)

Edellä lueteltuja tasoja (0-1B) ei jaeta avoimesti käyttäjille, kun taas avointa aineistoa ovat seuraavan prosessoinnin 1C-tason tuotteet ja niitä voi ladata esimerkiksi osoitteesta: <https://cophub.copernicus.eu/>. (ESA Sentinel Online 2017.) Tasolle 1C siirryttäessä suoritetaan geometrinen korjaus, jossa tavoitteena on tehdä kaikille kuville ja kohteille sekä ajallinen että spektrinen rekisteröinti. Tämän tavoitteen saavuttamiseksi käytetään 1A- ja 1B-tasojen välissä kehitettyä geometristä mallia, jota sovelletaan kaikille kanaville. (Gascon et al. 2014.) Geometrisen korjausprosessoinnin lopputuotteena syntyy 1C-tason orto-oikaistu kuva UTM/WGS84-projektiossa, jonka heijastussuhde vastaa ilmakehän yläpuolelta mitattua arvoa (engl. Top-Of-Atmosphere, TOA). Tason 1C tuotteet jaetaan ennalta määrätyn kokoisissa tiilissä (engl. tile), joiden koko on  $100 \times 100 \text{ km}^2$  ja se myös sisältää pilvi-, maa- ja vesimaskit. (Gatti ja Naud 2017.)

Seuraavien tasojen Sentinel-2-aineiston prosessoinnit suoritetaan käyttäjän toimesta Sentinel-2-työkaluilla. Sentinel-2-työkaluihin kuuluu välineitä, jotka sopivat visualisointiin, analysointiin ja prosessointiin. (ESA Sentinel Online 2017.) Esimerkiksi näistä työkaluista löytyy Sen2Cor-prosessori, jolla voidaan suorittaa ilmakehäkorjaus 1C-tason tuotteelle. Tällöin syntyy 2A-tason tuote, jonka arvot vastaavat maanpinnalta mitattua heijastusta (engl. Bottom-Of-Atmosphere, BOA). Tason-2A tuotteita on tarjolla eri spatiaalisilla resoluutiolla (10 m, 20 m ja 60 m) ja lisäksi syntyy aerosolien optista paksuutta (engl. Aerosol Optical Thickness, AOT), vesihöyryä (engl. Water Vapour, WV) sekä näkymän luokittelua (engl. Scene Classification, SC) kuvaavat karttatuotteet. (Gascon et al. 2017.)

Sen2Cor-ilmakehäkorjaus (Kuva 5) alkaa pilvitunnistuksella ja luokittelulla pilvisyyttä, maanpeitettä, vesistöä ja virheellisiä pikseleitä kuvaaviin luokkiin (yhteensä 11 luokkaa) (Main-Knorn et al. 2017). Tämän jälkeen otetaan huomioon aerosolien optinen paksuus sekä vesihöyrypitoisuus, jotka ovat määritetty Sentinel-2-kuvan kanavien avulla (Gascon et al. 2014). Viimeinen vaihe on muunnos TOA-heijastussuhteesta BOA-heijastussuhteeseen (Gascon et al. 2017).





Kuva 5. Sen2Cor-ilmakehäkorjauksen prosessointivaiheet tasojen 1C ja 2A välillä. (mukailen Gascon et al. 2017)

Sen2Cor-ilmakehäkorjaus on riippuvainen kahdesta lisäaineistosta, joista ensimmäinen on hakutaulukko (LUT) (Gascon et al. 2017), joka perustuu säteilyn kulkua kuvaavaan malliin libRadtran (Mayer ja Kylling 2005). Hakutaulukossa huomioidaan erilaiset ilmakehän olosuhteet, aurinkogeometrit sekä maanpinnan korkeudet (Gascon et al. 2014). Toinen lisäaineisto on digitaalinen korkeusmalli (engl. digital elevation model, DEM), joka tulee 1C-tason tuotteen mukana oletuksenaan 90 m spatiaalinen resoluutio. Halutessaan käyttäjä voi vaihtaa Sen2Cor-prosessointiin muun DTED-tallennusmuodossa olevan korkeusmallin. (Gascon et al. 2017.)

### 2.3 Metsän mittaus maastossa

Metsävara-arvioinnin peruselementti on yksittäinen puu, joka muodostuu maanpinnan alapuoleisista osista kuten juuret ja yläpuoleisista osista kuten runko, kuori, kanto, oksat sekä lehdet tai neulaset (Gschwantner et al. 2009). Metsän mittauksen näkökulmasta näistä osista kiinnostavin puun osa on runko ja erityisesti sen tilavuus (Kilkki 1983). Rungon tilavuus tarkoittaa rungon kuorellista tilavuutta kannon korkeudesta latvan huippuun (Lindblad 2014). Elävän pystypuun tilavuutta ei voida suoraan mitata, vaan se joudutaan johtamaan muiden tunnusten avulla. Yksittäistä puuta kuvaavia tunnuksia tilavuuden v lisäksi on rinnankorkeusläpimitta  $d_{1,3}$  (mitataan 1,3 m korkeudelta maanpinnasta), pituus  $h$ , pohjapinta-ala  $g$  sekä rinnankorkeusmuotoluku  $f$ , joka kuvaa puun muotoa (Taulukko 2). (Laasasenaho 1982.)

Taulukko 2. Yksittäistä puuta kuvaavia tunnuksia.

Tunnus	Lyhenne	Yksikkö
Rinnankorkeusläpimitta	$d_{1,3}$	cm
Pohjapinta-ala	$g$	m <sup>2</sup>
Pituus	$h$	m
Tilavuus	$v$	m <sup>3</sup>
Rinnankorkeusmuotoluku	$f$	-



Puunrunгон eri osat muistuttavat erilaisia geometrisia kappaleita, joita voidaan hyödyntää tilavuusarvioinnissa. Esimerkiksi puun tyveä voidaan kuvata neiloidilla, keskiosia paraboloidilla ja puun latvaa kartiolla tai paraboloidilla. Näiden geometrinen pyörähdyskappaleiden tilavuuskaavojen avulla voidaan arvioida puun tilavuutta. (Laasasenaho 1982.) Tilavuusmallin tarkkuuteen vaikuttaa maaston mittausvirheet sekä tilavuusmallissa käytettyjen muiden mallien virheet. Keskivirheellä kuvataan yksittäisten runkojen tilavuuksien hajaantumista laskettujen tulosten ympärille. Tilavuuden keskivirhe kasvaa maastossa mitattujen läpimitan ja pituuden virheiden vuoksi. Toisaalta jos pituus on arvioitu pituusmallilla maastomittauksien sijaan, niin silloin pituusmallin virhe lisää myös tilavuusmallin virhettä. Esimerkiksi jos pituusmalli antaa yliarvioitun tuloksen, niin silloin tilavuusmallin tulos on myös todennäköisesti yliarvio. (Päivinen et al. 1992.)

Suurta metsäaluetta arvioidessa yksittäisten puiden mittaaminen on aikaa vievää ja kallista koko puuston määrän selvittämiseksi, minkä seurauksena arviointi suoritetaan kuvioittain tai koealamittauksina (Poso 1983). Suomessa metsikkökuviot ovat kehitysluokaltaan, iältään, puulajijakaumaltaan ja muilta ominaisuuksiltaan mahdollisimman homogeenisiä metsäalueita, joiden koko on noin 1-20 ha (Mäkelä ja Pekkarinen 2004).

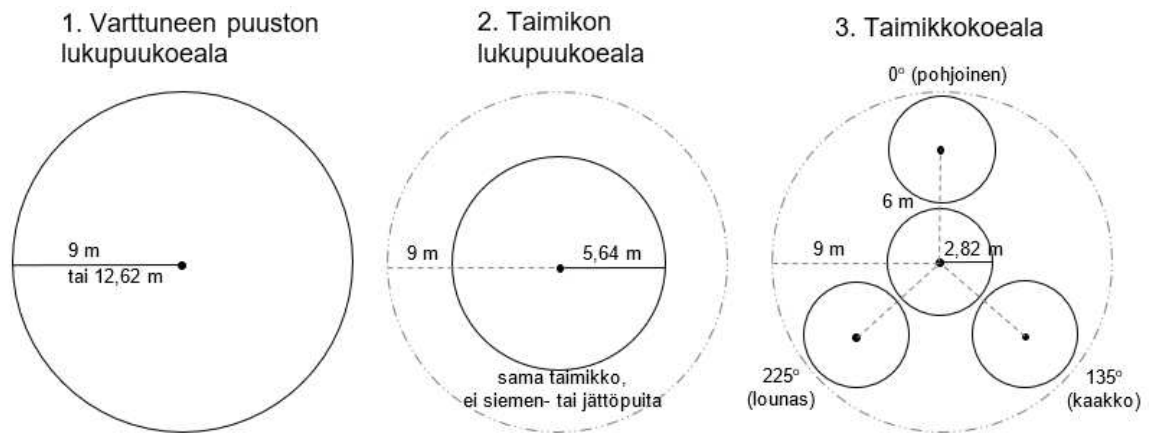
Metsikön tilavuuden arvioimiseksi on erilaisia toimintatapoja. Ensimmäinen arviointitapa on silmävarainen arviointi. Tämä tapa perustuu metsäasiantuntijan kokemukseen erilaisten tilavuuksien arvioimiseksi, jolloin tulos pohjautuu täysin havainnoijan taitoihin. Arvioinnin tarkkuutta on vaikea määrittää ja sen seurauksena sitä ei voida käyttää tieteellisissä, taloudellisissa tai lakisäätöisissä sovelluksissa. (Kershaw et al. 2017.)

Metsäalueen inventointi on toinen tapa, jossa metsikön tilavuutta arvioidaan koealojen avulla. Koeala, joka on pieni otos, edustaa koko arvioitavaa aluetta. Koealoja voidaan käyttää inventoinnin lisäksi myös metsän tutkimuksessa. (Poso 1983.) Esimerkiksi valtakunnallisella tasolla on suoritettu valtakunnan metsien inventointia (VMI). Suomessa ensimmäinen VMI toteutettiin vuosina 1921-1924 ja siitä lähtien inventointeja on toteutettu enemmän tai vähemmän säännöllisesti. VMI:ssa metsistä kerätään tietoa maastomittauksien avulla ja tyyppillisesti yhdeltä koealalta tallennetaan noin 200 muuttujaa, jotka muun muassa kuvaavat maaperää ja maankäyttöä sekä puun ja puuston ominaisuuksia. (Tomppo et al. 2008.)

Vuoteen 1989 asti VMI on perustunut ainoastaan maastomittauksiin (Tomppo ja Katila 1991). Uuden teknologian kuten satelliittikuvien myötä monilähteisen valtakunnan metsien inventoinnin (MVMI) kehitys on alkanut vuodesta 1989. Ensimmäiset MVMI:n tulokset laskettiin vuonna 1990. Monilähteisessä valtakunnan metsien inventoinnissa hyödynnetään satelliittikuvaa, maastomittauksia sekä muita tietolähteitä kuten digitaalisia karttoja. Lopputuotteena syntyy tilastoja pienistä alueista, estimaatteja kunnittain sekä rasterimuotoisia teemakarttoja, joiden spatiaalinen resoluutio on 25 m x 25 m. Teemakarttatason kuvaavat esimerkiksi metsikön ikää, pohjapinta-alaa ( $m^2$ ), kokonaistilavuutta hehtaaria kohden ( $m^3/ha$ ) sekä mänty-, kuusi- ja koivukohtaisia tilavuuksia hehtaaria kohden ( $m^3/ha$ ). (Tomppo et al. 2008.)

Maastossa mitattuja koealoja on eri muotoisia ja ne voivat olla kiinteä- tai muuttuva-alaisia. Kiinteäalaisia koealoja ovat metsikkökoeala sekä ympyräkoeala. Metsikkökoeala on suora-kaiteen muotoinen, kun taas ympyräkoeala on nimensä mukaisesti ympyrän muotoinen. Ympyräkoealasta mitataan tarkemmin tietyt puut eli lukupuut ja niiden säde vaihtelee sen mu-

kaan, onko kyseessä varttunut puusto vai taimikko (Kuva 6). Ympyräkoalat rajataan sädenauhan avulla, eikä niitä tarvitse merkitä maastoon. Tämän seurauksena ympyräkoalan mittaaminen on nopeampaa kuin metsikkökoalan. (Lindsey et al. 1958.)



Kuva 6. Kiinteäalalaisen koealan tyypit. (Metsäkeskus 2014)

Muuttuva-alainen koeala mittaa eri kokoisia puita eri suuruiselta alalta. Yksi muuttuva-alaisen koealan tyyppi on relaskooppikoeala. Koeala mitataan relaskoopilla, jossa varren ja hahlon avulla määritetään koealaan mukaan luettavat puut. Kaikki relaskooppikoealan puut edustavat yhtä suurta pohjapinta-alaa tiettyä pinta-alaa kohden. (Lehmonen 2014.)

Koealasta mitattujen tietojen ja mallien avulla ennustetaan tulokset koko metsikön alueelle. Koealan valintatapa ja mittaustarkkuus vaikuttavat siihen, kuinka hyvin se edustaa alueen puustoa. (Landsberg 2003.) Metsikön puustoa kuvaavia keskitunnuksia ovat esimerkiksi ikä  $T$ , keskiläpimitta  $D$  sekä -pituus  $H$ , kun taas hehtaarikohtaisia puustotunnuksia ovat esimerkiksi pohjapinta-ala  $G$ , tilavuus  $V$  (Nyyssönen ja Mielikäinen 1978) ja runkoluku  $N$  (Lehmonen 2014) (Taulukko 3). Runkolukua käytetään taimikoissa ja nuorissa metsissä pohjapinta-alan sijaan kuvaamaan puuston tiheyttä (Lehmonen 2014). Varttuneesta puustosta määritettävä pohjapinta-ala kuvaa muodostuvaa neliömetrimäärä, jos kaikista hehtaarin alueella kasvavista puista lasketut pinta-alat lasketaan yhteen rinnankorkeudelta. Yhden metsäalueen kaikkien puiden tilavuutta kutsutaan kokonaistilavuudeksi, kun taas keskitilavuus kuvaa yhden hehtaarin keskimääräistä puun määrää. Pystyssä olevan elävän puuston tilavuutta voidaan arvioida pohjapinta-alan avulla. (Metsäkeskus 2016.)

Taulukko 3. Metsikön puustoa kuvaavia tunnuksia.

Tunnus	Lyhenne	Yksikkö
Keskiläpimitta	$D$	cm
Pohjapinta-ala	$G$	m <sup>2</sup> /ha
Runkoluku	$N$	kpl/ha
Keskipituus	$H$	m
Puuston tilavuus	$V$	m <sup>3</sup> /ha
Ikä	$T$	a

Tilavuuden laskemiseksi voidaan käyttää erilaisia yhtälöitä. Pohjapinta-alaan, pituuteen sekä metsikön muotoa kuvaavaan lukuun perustuva tilavuusmalli kirjoitetaan alla esitetyllä yhtälöllä. (Kershaw et al. 2017.)

$$V = G \times H \times F \quad (1)$$

missä            V on puuston kokonaistilavuus [m<sup>3</sup>]  
                   G on puuston kokonaispohjapinta-ala [m<sup>2</sup>]  
                   H on keskipituus [m]  
                   F on metsikön muotoluku [-]

Tämän tilavuusmallin tuloksen tarkkuus perustuu yhtälössä esiintyvien muuttujien tarkkuuteen. Muotolukuun perustuva malli on karkea arvio puuston tilavuudesta, sillä puun muotoa kuvaava muotoluku on vaikea määrittää tarkasti. (Kershaw et al. 2017.) Tilavuusmalleja voidaan muodostaa pohjapinta-alan sekä keskipituuden lisäksi keskiläpimittaan perustuen. Laasasenaho (1982) määritti tutkimuksessaan rungon läpimittaa eri korkeuksilla kuvaavat runkokäyräyhtälöt. Näiden yhtälöiden avulla hän laati männyn, kuusen sekä koivun tilavuutta kuvaavat taulukot, jotka perustuvat puun rinnankorkeusläpimittaan sekä pituuteen (Laasasenaho ja Snellman 1983). Muita menetelmiä tilavuuden arvioimiseksi ovat esimerkiksi tiheysfunktio, joissa hyödynnetään beta- (Päivinen 1980) ja Weibull-jakaumia (Kilkki ja Päivinen 1986). Näitä suhteelliseen pohjapinta-alaan perustuvia jakaumia käytetään runkolukusarjan kuvaamiseen, joista edelleen lasketaan kokonaistilavuus. (Päivinen 1980, Kilkki ja Päivinen 1986.)

Metsän mittauksen sekä arvioinnin tavoitteena on määrittää kuviokohtainen runkotilavuus ja puutavaralajien määrä 20 % virhemarginaalilla (Lehmonen 2014). Yleensä virheet aiheutuvat esimerkiksi mittaajasta, mittalaitteesta, mittausolosuhteista sekä mitattavasta kohteesta (Kilkki 1983). Mittausvirheet voivat olla joko satunnaisia tai systemaattisia. Satunnaisen virheen vaikutus pienenee suuressa otosjoukossa. Suuremmalla otoskoolla ei kuitenkaan ole vaikutusta systemaattiseen virheeseen, joka voi syntyä esimerkiksi koealamittauksen aikana arvioijan lukuvirheestä. Esimerkiksi relaskoopin avulla tehtävissä mittauksissa eri henkilöiden arviot voivat vaihdella samalla kuviolla 15-20 %. Jos arvioitavia kuvioita on muutamia, niin silloin arvioinnin merkitys korostuu. Kuitenkin käytännössä eniten virhettä aiheutuu koealojen huonosta sijoittelusta, jolloin koealat eivät ole riittävän edustavia. (Paananen 2009.)

## **2.4 Puuston tilavuuden arvioiminen satelliittikuvalta**

### **2.4.1 Tilavuuden arviointimenetelmiä**

Puustotunnusten kuten runkotilavuuden, pohjapinta-alan ja latvuspeiton mittaaminen perustuen maastossa suoritettaviin mittausmenetelmin on kallista ja aikaa vievää. Ratkaisuna metsän arvioinnin ja mittauksen tehostamiseksi on hyödyntää kaukokartoitusaineistoa kuten satelliittikuvaa. Metsän inventoinnin näkökulmasta optinen satelliittikuva sisältää tutkakuvaa enemmän informaatiota (Hyypä et al. 2000).

Tilavuusmallin määrittämiseksi tulee ensin yhdistää maastosta mitatut referenssinä toimivat koealatiedot satelliittikuvan heijastussuhdearvoihin. Yhden pikselin sijaan joissain tapauksissa voi olla tarpeen käyttää myös ympäröivien pikselien heijastussuhdetietoa. Koealaa vastaava heijastussuhde voidaan esimerkiksi valita laskemalla heijastussuhteiden keskiarvo täydessä 3x3 ”kuvaikkunassa” (engl. image window) tai tämän lisäksi segmentoida arvot ikkunan sisällä. Segmentoinnissa jätetään ulkopuolelle merkitsevästi poikkeavat naapuripikselit. Segmentointi ikkunan sisällä voi tuottaa korkeampia korrelaatiokertoimia kuin käyttämällä kaikissa tilanteissa heijastussuhteen määrittämiseen täyttä 3x3-ikkunaa (Hall et al. 2006).

Selitettävänä muuttujana puuston tilavuusarvioinnissa on luonnollisesti puuston tilavuus. Kuitenkin menetelmän vaatiessa alkuperäisen mitatun puuston tilavuuden sijaan selitettävänä muuttujana voidaan käyttää myös tilavuuden luonnollista logaritmia. Esimerkiksi Halme ja Tomppo (2001) sekä Trotter et al. (1997) käyttivät tutkimuksissaan tilavuuden luonnollista logaritmia, jotta suhde muuttujien välillä oli lineaarinen.

Tilavuuden arvioinnissa mallin selittävinä muuttujina voidaan käyttää kanavien heijastussuhteiden lisäksi myös Haralick entropiaa (Astola et al. 2004) kasvillisuusindeksejä (Mohammadi et al. 2010) sekä metsikön tunnuksia kuten keskipituutta ja puuston ikää (Holmgren et al. 2000). Selittävien muuttujien valintaan vaikuttaa tilavuusarvioinnissa käytetty menetelmä. Puuston tilavuuden arvioimiseen käytettyjä menetelmiä ovat esimerkiksi k-NN-menetelmä (Mäkelä ja Pekkarinen 2004, Huiyan et al. 2006, Holmgren et al. 2000), k-MSN-menetelmä (Peuhkurinen et al. 2008), Forestime-ohjelmisto (Astola et al. 2004), BioSTRUCT-menetelmä (Hall et al. 2006), regressioanalyysi (Günlü et al. 2015, Mohammadi et al. 2010, Trotter et al. 1997, Hyypä et al. 2000), Random Forest -regressio (Chrysafis et al. 2017) sekä neuroverkkomalli (Hyypä et al. 2000). Tässä tutkimuksessa puuston tilavuuden arvioinnissa käytetään lineaarista regressioanalyysiä, joka selitetään tarkemmin luvussa 2.4.2.

k-NN-menetelmässä lähin naapuri eli tässä tapauksessa pikseli on määritelty tietyn etäisyyden perusteella. Etäisyydellä painotetussa k-NN-menetelmässä lähimpien naapurien vaikutus pienenee etäisyyden kasvaessa. (Huiyan et al. 2006.) k-MSN-menetelmä on k-NN-menetelmän sovellus, jossa käytetään tietynlaista etäisyysfunktiota. Etäisyysfunktio perustuu k-määrään lähimpiä naapureita, joilla on tietyt samankaltaiset ominaisuudet. (Peuhkurinen et al. 2008.) Etäisyys voidaan mitata esimerkiksi euklidisena (Mäkelä ja Pekkarinen 2004) tai Mahalanobisena etäisyytenä (Holmgren et al. 2000). Holmgren et al. (2000) käyttivät tilavuusarvioinnissa Mahalanobista etäisyyttä, sillä euklidinen etäisyys ei ota huomioon ominaisuuksien välisiä kovariansseja. Suomen valtakunnan metsien inventointi (VMI) perustuu k-NN-menetelmään, jossa käytetään painotettua euklidista etäisyyttä ja k:n arvo on 5-10 (Tomppo et al. 2008).

Etuna k-NN-menetelmässä on, että useita metsän tunnuksia voidaan arvioida samanaikaisesti ja että muuttujien väliset riippuvuussuhteet säilyvät (Moeur 1988). Mäkelä ja Pekkarinen (2004) pitivät tutkimuksensa perusteella k-NN-menetelmää sopivimpana alueille, joissa metsä on homogeenisempaa ja metsikköalueet ovat suurempia kuin esimerkiksi Etelä-Suomessa. Menetelmässä ilmenee heikkouksia, kun opetusaineiston koko on joko liian suuri tai pieni. Liian suuri opetusaineisto kasvattaa prosessointiaikaa (Tomppo et al. 2008). Toisaalta liian vähäisessä opetusaineistossa ei välttämättä ole k-kappaletta samanlaisia metsiköitä, jolloin tulos ei välttämättä edusta arvioitavaa aluetta (Mäkelä ja Pekkarinen 2004).

Forestime on Astolan et al. (2004) tutkimuksessa kehitetty ohjelmisto, joka soveltuu laajojen metsäalueiden kartoittamiseen automaattisesti korkearesoluutioiselta satelliittikuvalta. Ohjelmiston tuottamien metsämuuttujien arvioiminen perustuu kuvasta segmentoituihin heijastussuhdearvoiltaan homogeenisiin ”mikrometsiköihin”, metsikkökohtaisten ominaisuusvektoreiden laskemiseen ja maastotiedon yhdistämiseen ohjaamattomassa klusteroinnissa syntyneiden klustereiden kanssa. Estimaatit kuten runkotilavuus tai runkoluku tuotetaan painotettuina summina syötteenä olleiden esimerkkiluokkien todennäköisyyksistä. Yläraja tilavuusarvioille ohjelmistossa on noin 300 m<sup>3</sup>/ha.

Astolan et al. (2004) tutkimuksessa käytetty Forestime-ohjelmisto oli vielä osittain toiminnallisuuksiltaan puutteellinen, eikä näin ollen ollut täysin valmis operatiiviseksi työkaluksi puuston arviointiin. Ohjelmiston vahvuus on sen joustavuus ja muokattavuus, esimerkiksi sen voi helposti konfiguroida käyttämään useita eri satelliittiaineistoja. Esimerkkisovelluksessa käytettiin IKONOS-satelliittikuvaa, jonka spatiaalinen resoluutio on yhdestä neljään metriä. Tuloksena syntyneiden tilavuusarvioiden suhteellinen virhe oli melko pieni (nRMSE 37,4 %) verrattuna esimerkiksi runkolukuarviointiin, jonka virheet olivat suurempia (nRMSE 87,1 %).

BioSTRUCT-menetelmässä kaukokartoitetun kuva-aineiston avulla ensin määritetään metsän tunnuksia kuten puuston pituutta ja latvuspeittoa. Tämän jälkeen mallinnettuja puustotunnuksia käytetään tilavuuden arvioimisessa. Hall et al. (2006) käyttivät kaukokartoitusaineistona Landsat-satelliitin kuvaa. He totesivat, että verrattuna pelkästään Landsat-satelliitin kanavien heijastussuhteeseen perustuvaan tilavuusarvioon BioSTRUCT-menetelmä parantaa tuloksen tarkkuutta. Tutkimuksessa käytetyt maastoaineistot sisälsivät pääasiassa havupuuta, jolloin lehti- ja sekametsään kohdistuvaa tutkimusta tarvitaan ennen kuin menetelmä soveltuu laajemmin metsien inventointiin ja metsätalouden sovelluksiin.

Regressioanalyysissä tutkitaan selitettävän muuttujan ja selittävien muuttujien välistä suhdetta. Esimerkiksi puuston tilavuutta arvioidessa selitettävänä muuttajana on tilavuus ja selittävinä muuttujina eri kanavien heijastussuhteet. (Mohammadi et al. 2010.) Useimmiten korrelaatio tilavuuden sekä kanavien heijastussuhteen välillä on negatiivinen (Trotter et al. 1997, Mohammadi et al. 2010, Chrysafis et al. 2017). Negatiivinen korrelaatio tarkoittaa tilavuuden kasvua heijastussuhteen pienentyessä.

Trotter et al. (1997) havaitsivat, että selittävinä muuttujinaan pelkästään kanavien heijastussuhteita käyttävän regressiomallin tulos ei parantunut kasvillisuusindeksien lisäyksellä. Tutkimuksen parhaimmat tilavuusmallit perustuivat Landsat-TM (Thematic Mapper) -satelliitin punaisen sekä NIR-aallonpituuden kanaviin. Mohammadi et al. (2010) vuorostaan havaitsivat Landsat ETM+ (Enhanced Thematic Mapper Plus) -satelliittikuvaa käyttävässä tutkimuksessaan, että vihreys (engl. greenness) ja DVI (Difference Vegetation Index) -kasvillisuusindeksi toimivat tilavuusarvioinnissa paremmin kuin mikään muu Landsat ETM+-kanava tai kasvillisuusindeksi. DVI-indeksi on NIR- ja SWIR- aallonpituuksilla sijaitsevien kanavien välinen erotus. Eroja selittävien muuttujien välillä selittää muun muassa tutkimuskohteiden maantieteellinen sijainti, puulajit ja referenssinä käytetty puuston tilavuusjakauma. Trotterin et al. (1997) tutkimuksessa käytetty kohde koostui havupuusta, sijaitsi Uudessa-Seelannissa ja referenssiaineiston keskitilavuus oli 413 m<sup>3</sup>/ha sekä minimi- ja maksimitilavuudet olivat 190 ja 810 m<sup>3</sup>/ha. Mohammadin et al. (2010) tutkimuskohde koostui pääosin lehtipuusta, joka sijaitti Iranissa. Verrattuna Trotterin et al. (1997) tutkimukseen Mohammadin et al. (2010) tutkimusmetsiköt olivat tilavuudeltaan pienempiä noin 50-550 m<sup>3</sup>/ha.

Random Forest -regressio on koneoppimismenetelmä, jossa muodostetaan annetun opetusaineiston avulla päätöspuita (Breiman 2001). Menetelmän vahvuutena on sen tehokkuus ja kyky käsitellä heikkojakin havaintoja (Breiman 2001), mutta samalla pitää huolehtia, ettei tulos ole aineistoon ylisovittunut (Grömping 2012). Chrysafis et al. (2017) selvittivät Random Forest -regressiolla kanavien heijastussuhteiden sekä puuston tilavuuden välistä suhdetta tutkimuskohteessaan Kreikassa, jossa hallitsevina puulajeina olivat mustamänty,

tammi sekä pyökki. Tutkimuksessa käytettiin sekä Sentinel-2- että Landsat 8-satelliittien aineistoa. Merkittävä ero näiden satelliittien välillä on Sentinel-2-satelliitin punaisen reunan alueen kanavat (B5, B6, B7), joita Landsat 8-satelliitissa ei ole. Random Forest -regressiomalli perustuen Sentinel-2-satelliitin kanaviin, joiden spatiaalinen resoluutio on 10 m (B2, B3, B4, B8) tai 20 m (B5, B6, B7, B8a, B11, B12) tuotti Chrysafisin et al. (2017) tutkimuksessa hiukan paremman tilavuusarvion kuin Landsat 8-satelliitin kanaviin perustuva malli. Sentinel-2-satelliitin kanaviin perustuvassa mallissa merkittävimmät kanavat olivat punaisen reunan (B5, B6) sekä SWIR-aallonpituuden (B11) alueilla. Lisäksi havaittiin, että kasvillisuusindeksit eivät toimineet paremmin verrattuna pelkkiin Sentinel-2-kanaviin.

Neuroverkkomalli on tekoälyyn perustuva menetelmä, jossa ensimmäisen joukon elementit järjestellään toisen joukon elementeiksi. Neuroverkkomalli kuuluu samaan menetelmäluokkaan regressioanalyysin kanssa. (Atkinson ja Tatnall 1997.) Hyypä et al. (2000) vertailivat neuroverkkomallin ja regressioanalyysin sekä erilaisien kaukokartoitusaineistojen soveltuvuutta keskipituuden, pohjapinta-alan ja runkotilavuuden arvioinnissa. Tutkimuksessa käytetty kaukokartoitusaineisto oli sekä kuvaavan spektrometrin aineistoa, ilma-, tutka- että optista satelliittikuvaa. Kuvaavan spektrometrin aineisto ja ilmakehä tuottivat parhaimmat arviot keskipituudesta, pohjapinta-alasta sekä runkotilavuudesta. Heikoimman arvion tuotti tutkakuvaa-aineisto, kun taas satelliittikuvien tuottamat arviot jäivät tutkakuvien ja ilmakehien tulosten väliin. Hyypä et al. (2000) havaitsivat, että valitulla menetelmällä (regressioanalyysi tai neuroverkkomalli) ei ollut merkittävää vaikutusta kaukokartoitusaineistosta arvioidun metsätunnuksen suuruuteen. Menetelmissä käytettyjä selittäviä muuttujia ei tutkimuksessa listata, sillä erilaisilla selittävien muuttujien yhdistelmillä saavutettiin samankaltaisia arviointilopputuloksia.

Eri menetelmien tuottamia tuloksia voidaan arvioida ja vertailla keskenään tilastollisilla analyysillä sekä mallin toimivuutta kuvaavilla tunnusluvuilla kuten korjattu  $R^2$ -luku, keskineliövirheen neliöjuuri (engl. root mean square error, RMSE) sekä normalisoidun keskineliövirheen neliöjuuri (nRMSE). Luvuissa 2.4.2 ja 3.3 käsitellään tarkemmin edellä mainittu tunnusluvut.

Puuston tilavuusarvioinnin tulokset edellä mainituissa tutkimuksissa vaihtelevat keskimäärin normalisoidulla keskineliövirheen neliöjuurella (nRMSE) 31,3 % (Peuhkurinen et al. 2008) ja 56 % (Hyypä et al. 2000) välillä. Keskineliövirheen neliöjuuren (RMSE) arvot vaihtelevat vuorostaan 52,2 m<sup>3</sup>/ha (Peuhkurinen et al. 2008) ja 110,8 m<sup>3</sup>/ha (Hall et al. 2006) välillä. Günlü et al. (2015) saivat edellä mainituista tutkimuksista poikkeavasti nRMSE-arvoksi 6,35 % ja RMSE-arvoksi 8,93 m<sup>3</sup>/ha. Tutkimuksessa tilavuuden arvioinnin menetelmänä oli regressioanalyysi, jossa selittävinä muuttujina käytettiin SPOT-4-satelliitin kanavia sekä niistä laskettuja kasvillisuusindeksejä. Paras tulos tilavuuden arvioimiseksi saavutettiin vihreän ja punaisen aallonpituuden kanavilla sekä SR- (Simple Ratio), DVI- (Difference Vegetation Index), SAVI- (Soil-Adjusted Vegetation Index), EVI- (Enhanced Vegetation Index) ja IPVI- (Infrared Percentage Vegetation Index) kasvillisuusindekseillä. Monet näistä indekseistä perustuvat punaisen sekä NIR-aallonpituuksien kanaviin, jolloin muuttujien välillä esiintyy todennäköisesti multikollineaarisuutta. Multikollineaarisuus aiheuttaa ongelmia regressiomalliin, joka voi osittain selittää Günlün et al. (2015) tutkimuksen poikkeavaa tulosta.

Huiyanin et al. (2006) sekä Mäkelän ja Pekkarisen (2004) tutkimuksien perusteella kokonaistilavuuden arvioiminen on tarkempaa kuin yksittäisten puulajien tilavuuden arviointi.

Lisäksi tutkimuksissa havaittiin, että lehtipuun tilavuuden arvioiminen on havupuita haasteellisempaa. Huiyan et al. (2006) tutkimuksessa kokonaistilavuuden nRMSE oli 44,2 %, havupuulajeista lehtikuusella 51,7 % ja Korean männyllä 71,7 % sekä lehtipuulla 88,2 %. Vastaavasti Mäkelän ja Pekkarisen (2004) tutkimuksessa kokonaistilavuuden nRMSE oli 48 %, havupuulajeista kuusella 81 % ja männyllä 117 % sekä lehtipuulla 174 %.

Useimpien tilavuusmallien haasteena on niiden tuottamien arvioiden tarkkuus. Metsän inventoinnissa tilavuusarvion virheen suuruus saa kuviokohtaisesti olla korkeintaan 20 % (Lehmonen 2014), jonka seurauksena pelkästään satelliittikuvaan pohjautuvat arviot ovat liian epätarkkoja metsätalouden tarpeisiin. Näin ollen tarvitaan uusia menetelmiä ja satelliittiaineistoja, jotta satelliittikuvaan pohjautuva tilavuusarvio on riittävän tarkka.

Edellä mainituissa tutkimuksissa ja menetelmissä lukuun ottamatta Chrysafisin et al. (2017) tutkimusta käytetyt satelliittien sensorit kuvaavat näkyvän valon, NIR- ja SWIR-aallonpituuksien alueella. Hyypä et al. (2000) totesivat, että metsäsovelluksiin tarvitaan instrumentteja, joiden kanavat ovat aallonpituusalueeltaan laajempia ja sijaitsevat spektrin eri alueilla. Tähän ongelmaan vastaa kolme Sentinel-2-kanavaa punaisen reunan aallonpituuden alueella. Chrysafis et al. (2017) osoittivat tutkimuksessaan näiden kanavien potentiaalin Välimeren metsien tilavuuden arvioinnissa. Tämä antaa viitteitä mahdollisista höydyistä käyttäen Sentinel-2-kanavia, erityisesti punaisen reunan alueella, myös pohjoisen havumetsävyöhykkeen metsissä.

## 2.4.2 Regressioanalyysi

Tässä tutkimuksessa tilavuuden arvioimiseksi valittu menetelmä on regressioanalyysi. Regressioanalyysi on useissa sovelluksissa yleisesti käytetty tilastollinen menetelmä, jonka avulla voidaan helposti mallintaa ja vertailla muuttujien välisiä riippuvuussuhteita (Yan ja Su 2009). Menetelmää on myös käytetty aikaisemmissa satelliittiaineistoon perustuvissa puustoa arvioivissa tutkimuksissa (Trotter et al. 1997, Günlü et al. 2015, Hyypä et al. 2000, Mohammadi et al. 2010).

Yksinkertaisessa lineaarisessa regressiomallissa selittävä muuttuja  $x$  on lineaarisessa suhteessa selitettävän muuttujan  $y$  kanssa. Esimerkiksi puuston tilavuutta mallinnetaan yhden kanavan heijastussuhteen avulla. Yksinkertainen lineaarinen regressiomalli voidaan kuvata alla esitetyllä yhtälöllä. (Montgomery et al. 2012.)

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2)$$

missä  $y$  on selitettävä muuttuja  
 $\beta_0$  on regressiokerroin  
 $\beta_1$  on regressiokerroin  
 $x$  on selittävä muuttuja  
 $\varepsilon$  on satunnaisvirhe

Usean muuttujan lineaarisessa regressioanalyysissä selittäviä muuttujia on yhden sijaan useita. Selitettävän muuttujan  $y$  suhde selittäviin muuttujiin esitetään seuraavasti (Montgomery et al. 2012)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (3)$$

missä  $y$  on selitettävä muuttuja  
 $\beta_j$  on regressiokerroin, jossa  $j = 0, 1, \dots, k$   
 $x_j$  on selittävä muuttuja, jossa  $j = 0, 1, \dots, k$   
 $\varepsilon$  on satunnaisvirhe

Mallin parametrien arvioimiseksi voidaan käyttää pienimmän neliösumman menetelmää (PNS-menetelmä). PNS-menetelmässä minimoidaan jäännösvirheiden neliösummaa regressiokertoimien estimaattien löytämiseksi. Näin löydettyt estimaatit muodostavat regressiomallin sovituksen, joka kuvaa selitettävän muuttujan  $y$  vaihtelua. (Yan ja Su 2009.)

Mallin soveltuvuutta voi arvioida varianssianalyysin (engl. analysis of variance, ANOVA) avulla. ANOVA perustuu muuttujan  $y$  kokonaisvaihtelun osittamiseen. Varianssianalyysin tulos kertoo, kuinka paljon muuttujan  $y$  vaihtelusta selittyy mallin avulla ja kuinka paljon jää selittämättä. Kokonaispoikkeamien neliöiden summa (SST) jakaantuu mallin selittämien poikkeamien summaan (SSR) ja selittämättömien poikkeamien summaan (SSE). (Montgomery et al. 2012, Yan ja Su 2009.)

Muita tapoja mallin soveltuvuuden arvioimiseen ovat  $R^2$ -luku ja korjattu  $R^2$ -luku.  $R^2$ -luvulla kuvataan, kuinka hyvin regressiomalli selittää muuttujan  $y$  vaihtelua. Lukuarvo vaihtelee 0 ja 1 välillä. Malli selittää suurimman osan muuttujan  $y$  vaihtelusta, kun  $R^2$ -luvun arvo lähestyy lukua 1. Muuttujien määrää lisäämällä myös  $R^2$ -luvun arvo kasvaa huolimatta siitä, onko lisätyllä selittävällä muuttujalla todellista vaikutusta selitettävään muuttujaan. Tämän seurauksena eri regressioanalyysien tuloksien vertailussa käytetään korjattua  $R^2$ -lukua. Korjattu  $R^2$ -luku huomioi muuttujien lukumäärän, jolloin se on joko yhtä suuri tai pienempi kuin  $R^2$ -luku. (Montgomery et al. 2012.)

Lineaariseen regressiomalliin liittyy oletuksia, jotka koskevat lineaarisuutta, riippumattomuutta, homoskedastisuutta sekä normaalisuutta. Lineaarioletuksen mukaan regressioanalyysillä voidaan tutkia muuttujien välisiä lineaarisia suhteita. Jos lineaarista yhteyttä ei esiinny, muuttujien välillä voi olla epälineaarinen yhteys. Riippumattomuus tarkoittaa, että satunnaisvirheet ovat toisistaan riippumattomia. Homoskedastisuutta koskeva oletus edellyttää, että kaikilla satunnaisvirheillä on yhtä suuri vakiona pysyvä varianssi. Normaalisuudella oletetaan normaalisti jakautuneita satunnaisvirheitä. (Yan ja Su 2009.)

Regressioanalyysin rajoitteena on selittävien muuttujien välillä esiintyvä multikollineaarisuus. Multikollineaarisuus tarkoittaa, että selittävät muuttujat korreloivat voimakkaasti keskenään. Jos muuttujien välinen korrelaatio on suuri, niin silloin regressioanalyysin tuloksien tarkkuudessa voi ilmetä ongelmia. Multikollineaarisuuden suuruutta voidaan arvioida VIF-mittarilla (engl. variance inflation factor). Pienin mahdollinen VIF-luku on 1. Multikollineaarisuudesta ja siitä aiheutuvista ongelmista viittaa VIF-luku, joka on yli 5. (Daoud 2017.)

Tämän tutkimuksen regressiomalli on kehitetty uusien havaintojen arvioimiseen ja näin ollen huolellinen validointi on tärkeää. Regressiomallin validoinnilla on tarkoitus selvittää mallin toimivuutta tarkoituksenmukaisessa ympäristössä. Validoinnissa voidaan käyttää seuraavia menetelmiä:

1. Mallin kertoimien ja ennustettujen arvojen analyysi
2. Mallin toimivuuden tutkiminen uudella aineistolla
3. Aineiston jakaminen eli ristiinvalidointi (Montgomery et al. 2012).



Mallin kertoimien ja ennustettujen arvojen analyysissä tutkitaan korrelaatiokertoimien vakautta, etumerkkiä sekä suuruutta. Aikaisempi kokemus, teoria tai analyttinen malli antavat suuntaa korrelaatiokertoimien suuruudesta ja vaikutuksista. Mallin kertoimia tarkastelemalla voidaan tunnistaa odottamattomat etumerkit tai liian suuret absoluuttiset arvot, jotka viittaavat mallin virheellisyyteen. (Montgomery et al. 2012.) Myös multikollineaarisuudesta kertova VIF-luku antaa viitteitä mallin validiteetista (Daoud 2017).

Toinen menetelmä mallin validoimiseksi vaatii uuden aineiston keräämistä. Mallia testataan uudella aineistolla ja saatuja arvioita verrataan mitattuihin havaintoihin. Jos vaihtoehtoisia malleja on useampia, niin uuden aineiston avulla voidaan suorittaa lopullisen mallin valinta. Tämän validointimenetelmän toteuttamisessa usein ongelmaksi muodostuu uuden aineiston rajoitettu saatavuus. Aineiston jakaminen tarjoaa ratkaisun, kun uuden aineiston kerääminen on mahdotonta. (Montgomery et al. 2012.)

Aineiston jakamista hyödyntävää validointimenetelmää kutsutaan myös ristiinvalidoinniksi. Ristiinvalidoinnissa aineisto jaetaan kahteen osaan: opetus- ja testiaineistoon. Opetusaineistoa käytetään mallin rakentamiseen, jonka jälkeen mallin toimivuutta tutkitaan testiaineiston avulla. Tupla-ristiinvalidointi (engl. double-cross validation) on ristiinvalidoinnin menetelmä, jossa opetus- ja testiaineiston paikkaa vaihdetaan ensimmäisen mallin rakentamisen jälkeen. Lopputuloksena on kolme mallia, joista yksi on muodostettu koko aineiston avulla ja kaksi muuta ovat syntyneet aineiston jakamisen seurauksena. Tupla-ristiinvalidoinnin etuna on, että se tarjoaa useamman arvion mallin toimivuudesta. (Montgomery et al. 2012.)



Taulukossa 5 esitetään luokitellun aineiston tunnuslukuja. Männyn osuus koko koeala-aineistosta oli 26,6 %, kuusen 27,4 % ja lehtimetsän 19,0 %. Havusekametsän että sekametsän osuus oli 8,8 % ja 18,2 %. Koko aineistolla puuston tilavuus oli välillä 0 - 924 m<sup>3</sup>/ha. Keskitilavuus oli mediaania korkeampi kaikilla puulajiluokilla, mikä viittaa oikealle vinoon jakaumaan.

*Taulukko 5. Koeala-aineiston tunnuslukuja tilavuudesta puulajiluokittain.*

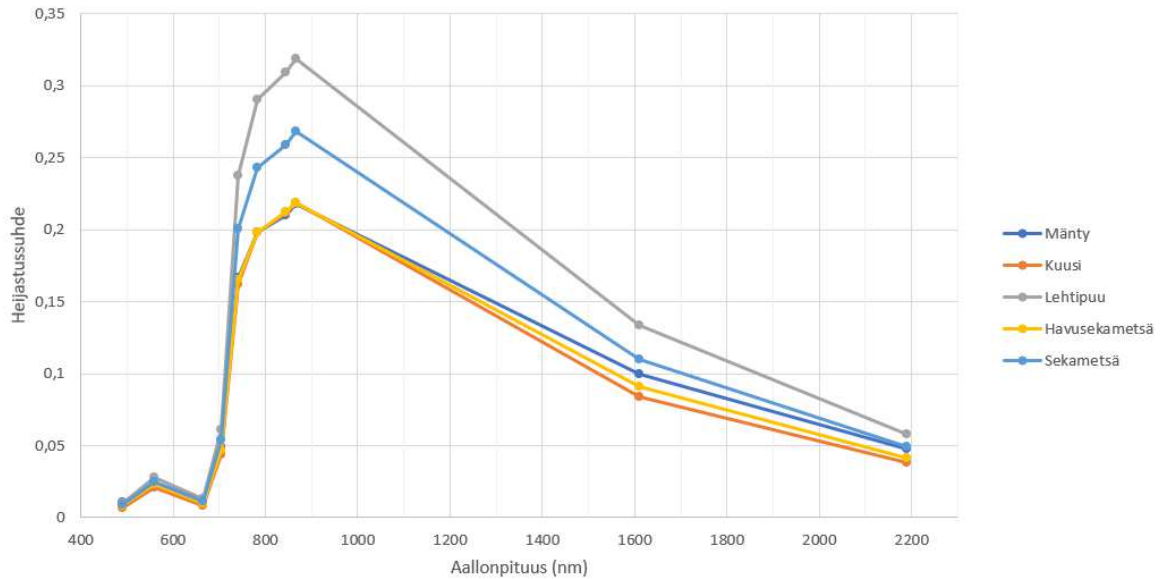
Puulajiluokka	Prosenttiosuus koko aineis- tosta (%)	Keskitila- vuus (m <sup>3</sup> /ha)	Mediaani tilavuus (m <sup>3</sup> /ha)	Tilavuuden keskihajonta (m <sup>3</sup> /ha)	Max tila- vuus (m <sup>3</sup> /ha)	Min tila- vuus (m <sup>3</sup> /ha)
Mänty	26,6	151,7	146,7	93,1	514,3	0,3
Kuusi	27,4	219,8	200,6	170,0	924,0	0,5
Lehtipuu	19,0	94,6	68,0	106,7	728,1	0,9
Havusekametsä	8,8	190,8	175,2	134,0	438,5	1,2
Sekametsä	18,2	148,2	109,4	140,3	573,2	0,0

### 3.2 Sentinel-2-satelliittikuva

Sentinel-2A-kuvien tarkastelu rajattiin 2015 vuoden kasvukauteen, jotta satelliittikuva ja koealamittaukset vastasivat mahdollisimman hyvin toisiaan. Valittu kuva oli vaihtoehtoista pilvettömin (pilvipeitto 0.036%) ja se oli otettu 17.8.2015 klo 9:50:11. Koealat sijoittuvat neljän tiilen alueelle T35VLH, T35VLG, T35VMG ja T35VMH, joiden auringonkulmat olivat vastaavassa järjestyksessä 47,7°, 46,9°, 46,7° ja 47,6°. Kuva on ladattu Copernicus Open Access Hub -sivuston kautta (<https://scihub.copernicus.eu/dhus/>). Kuvan prosessointitaso oli 1C, joka tarkoittaa orto-oikaistuja kuvia UTM/WGS84 projektiossa. 1C-tason pikselien arvot vastasivat TOA heijastussuhdetta.

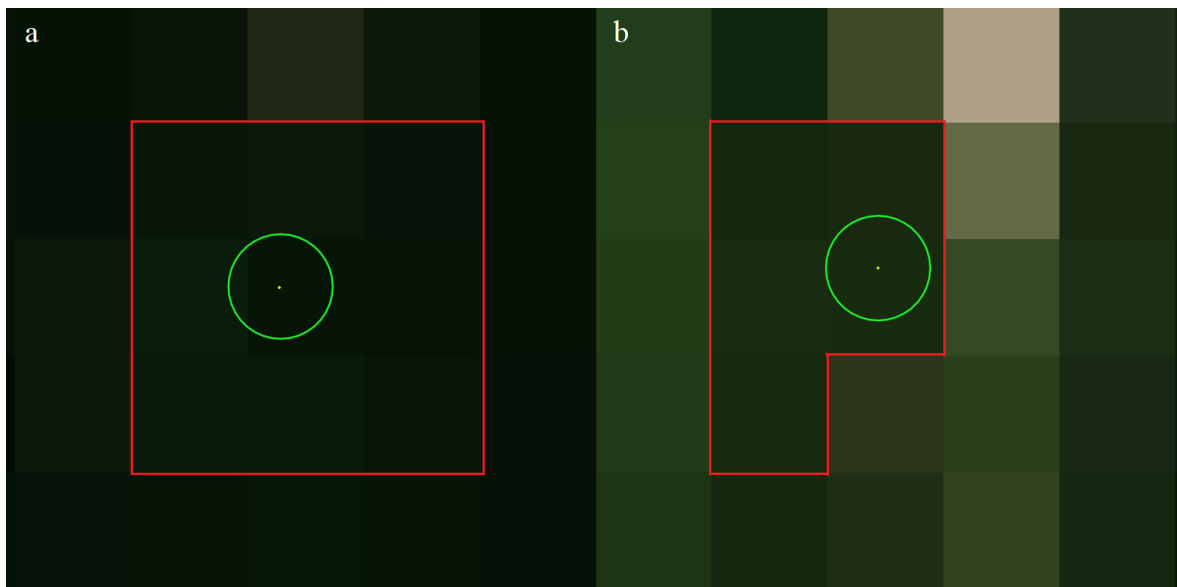
Kuvalle suoritettiin ilmakehäkorjaus Sentinel-2 Toolbox:sta löytyvän Sen2Cor (moduuliversio 2.4.0) avulla. Sen2Cor ajettiin komentoriviltä, jolle annettiin syötteenä 1C-tason kuva. Algoritmi käyttää korjauksessa hyväksi ilmakehän aerosoleja, vesihöyryä ja cirrusta kuvaavia kanavia B1, B9 ja B10 ja muuttaa pikselien arvot TOA heijastussuhteesta vastaamaan BOA heijastussuhdetta. Tämän tutkimuksen tilavuusarvioinnissa ei tarkasteltu ilmakehää kuvaavia kanavia, sillä niiden spatiaalinen resoluutio on 60 m ja pääasiallinen käyttötarkoitus on ilmakehän kuvaamisessa.

Koeala-aineiston eri puulajiluokat erottuivat toisistaan heijastusspektrin avulla (Kuva 8). Parhaiten erottuivat lehtipuu sekä sekametsä punaisen reunan (B5, B6, B7), NIR- (B8, B8a) ja SWIR-aallonpituusalueilla (B11). Mänty, kuusi sekä havusekametsä eivät yhtä selkeästi erottuneet toisistaan muilla kuin SWIR-aallonpituusalueilla (B11, B12). Tämä on linjassa Hovi et al. (2017) tutkimuksen kanssa, jossa he havaitsivat havupuiden erojen kasvavan SWIR-aallonpituusalueita lähestyttäessä.



Kuva 8. Puulajiluokkien keskiarvoinen heijastusspektri. Heijastusspektrin taitepisteet vastaavat Sentinel-2-kanavia B2, B3, B4, B5, B6, B7, B8, B8a, B11 ja B12. Koealojen lukumäärä eriluokilla: mänty = 218, kuusi = 225, lehtipuu = 156, havusekametsä = 72 ja sekametsä = 150.

Koealaa ja sitä vastaavan eri kanavien heijastusarvon yhdistämisessä otettiin huomioon keskipistettä vastaavan pikselin lisäksi 8 naapuripikseliä, jolloin muodostui 3x3 pikselin kokoinen ”kuvaikkuna” (engl. image window). Esimerkiksi havumetsillä, joita Suomen metsät pääasiassa ovat, useamman pikselin käyttö yhden pikselin sijaan vähentää tilavuuden arvioinnin virhettä. Maastossa sijaitsevan koealan tarkka paikannus kuvalla on vaikeaa ja lisäksi kuvan geometrisessa oikaisussa pikselien sijainnit voivat vääristyä (Halme ja Tomppo 2001). Myöskään mitattu koeala ei sijainnut täsmällisesti yhden pikselin alueella (Kuva 9).

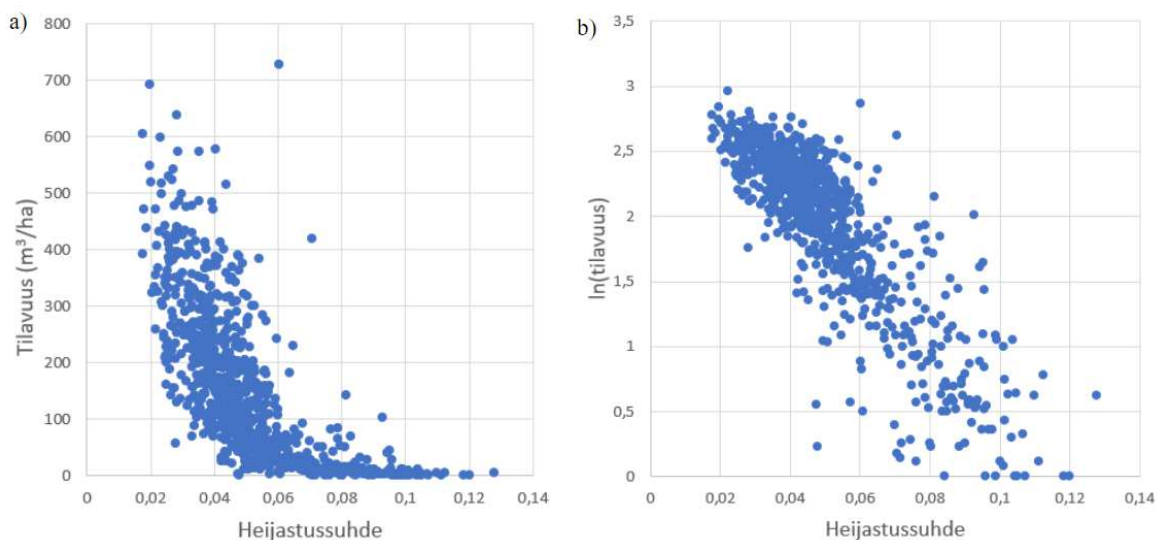


Kuva 9. Koealaa vastaavan heijastusarvon määrittäminen, kun (a) kaikki naapurit otetaan huomioon tai (b) naapurit, jotka poikkeavat keskipisteen pikselistä alle 20 %.

Naapuripikseleistä otettiin huomioon pikselit, joiden arvo poikkesi keskipisteen pikselin arvosta alle 20 % (Kuva 9b). Merkitsevien pikseleiden heijastusarvoista laskettiin keskiarvo, jota käytettiin koealaa vastaavana heijastusarvona. Osa koealoista oli lähellä tienreunaa tai metsäkuvioiden rajalla, jolloin heijastusarvoltaan poikkeavat pikselit todennäköisesti kuvasivat muuta kuin koealaa. Useassa tapauksessa koealan ympäristössä oli samankaltaista metsää, joten useamman pikselin tarkastelulla on mahdollista vähentää yksittäisen pikselin satunnaisvirheen vaikutusta.

### 3.3 Regressiomalli puuston tilavuuden ennustamiseksi

Mallin kehittämisessä tutkittiin selitettävänä muuttujana sekä tilavuutta että tilavuuden luonnollista logaritmia, sillä Halmeen ja Tompon (2001) tutkimuksessa kanavien ja tilavuuden luonnollisen logaritmin välillä esiintyi selkeä lineaarinen suhde. Kuvassa 10 havainnollistetaan punaisen reunan ensimmäisen kanavan (B5) ja tilavuuden sekä tilavuuden luonnollisen logaritmin välistä suhdetta. Kuvan 10 perusteella lineaarinen suhde oli voimakkaampi kanavan B5 ja tilavuuden luonnollisen logaritmin välillä.



Kuva 10. Yhteys kanavan B5 (keskiaallonpituus 703,8 nm) sekä a) tilavuuden että b) tilavuuden luonnollisen logaritmin välillä.

Selittävinä muuttujana tai selittävinä muuttujina olivat Sentinel-2-satelliitin kanavat B2, B3, B4, B5, B6, B7, B8, B8a, B11 ja B12. Tässä tutkimuksessa rajattiin ulkopuolelle kasvillisuusindeksit Chrysafisin et al. (2017) tutkimustulokseen perustuen. Heidän tutkimuksessa havaittiin, että kasvillisuusindeksit eivät parantaneet tilavuusarvioinnin tarkkuutta verrattuna pelkkiin Sentinel-2-kanaviin. Lineaariseen regressioanalyysiin soveltuvien kanavien määrittämisessä selvitettiin muuttujien väliset korrelaatiokertoimet. Pearsonin korrelaatiokertoimien laskentaan käytettiin IBM SPSS Statistics 24 -ohjelmistoa. Kertoimien tuloksien mukaan valittiin parhaat kanavat tai kanavayhdistelmät, joiden välillä ei esiintynyt voimakasta korrelaatiota ( $> 0,9$ ) multikollineaarisuuden välttämiseksi.

Lineaarinen regressioanalyysi suoritettiin puolajiluokittain selitettävän ja selittävän muuttujan välillä IBM SPSS Statistics 24 -ohjelmistolla. Jokaisen mallin tuli täyttää seuraavat kriteerit:

1. Mallin selittävien muuttujien VIF-luku ei saanut olla yli 5.
2. Kaikkien selittävien muuttujien tuli olla tilastollisesti merkittäviä ( $p < 0,05$ ).

3. Mallin regressiokertoimien tuli olla etumerkiltään samat kuin Pearsonin korrelaatio-  
kertoimien etumerkit selittävien ja selitettävien muuttujien välillä.

Näiden kolmen kriteerin avulla pyrittiin estämään virheellisen mallin syntyminen. Liian suuri VIF-luku, tilastollisesti merkitsemättömät muuttujat sekä mallissa vaihtelevat etumerkit viittaavat epävakaaseen regressiomalliin tai heikkoon arvioon yksittäisten regressiokertoimien vaikutuksista (Montgomery et al. 2012).

Tuloksena saatuja regressiomalleja vertailtiin puulajiluokittain käyttämällä korjattua  $R^2$  -lukua. Lisäksi laskettiin keskineliövirheen neliöjuuri (engl. root mean square error, RMSE), normalisoitu keskineliövirheen neliöjuuri (nRMSE), absoluuttinen mediaanivirhe (engl. median absolute error, MdAE) sekä normalisoitu absoluuttinen mediaanivirhe (nMdAE). MdAE ei ole yhtä herkkä poikkeaville havainnoille kuin RMSE. Edellä luetellut tunnusluvut lasketaan seuraavien yhtälöiden mukaisesti.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

missä  $n$  on havaintojen lukumäärä  
 $y_i$  on havaittu arvo  
 $\hat{y}_i$  on ennustettu arvo

normalisoitu RMSE

$$nRMSE = \frac{RMSE}{\bar{y}} \quad (5)$$

missä  $\bar{y}$  on havaintojen keskiarvo

absoluuttinen mediaanivirhe

$$MdAE = Md(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|) \quad (6)$$

normalisoitu absoluuttinen mediaanivirhe

$$nMdAE = \frac{MdAE}{\text{mediaani havainnoista}} \quad (7)$$

Lopullisen mallin valinta perustui korjatun  $R^2$  -luvun, RMSE:n, nRMSE:n, MdAE:n, nMdAE:n, mallin tuottaman maksimin sekä minimin vertailuun. Suurin painoarvo oli nRMSE:n ja nMdAE:n tuloksilla. Tavoitteena oli löytää malli, jonka  $R^2$ -luku oli mahdollisimman suuri, mutta samalla RMSE:n, nRMSE:n, MdAE:n ja nMdAE:n tulos mahdollisimman pieni.

Mallien validointiin käytettiin tupla-ristiinvalidointimenetelmää, jossa ensin laskettiin lineaarinen regressiomalli käyttäen kaikkia puulajiluokan havaintoja. Tämän jälkeen aineisto jaettiin satunnaisesti kahteen osaan, opetus- ja testiaineistoon. Opetusaineistolla laskettiin regressiomalli, jota testattiin testiaineistolla. Tämän jälkeen aineistojen paikkaa vaihdettiin ja

sama prosessi suoritettiin uudelleen. Tuloksena syntyi kolme mallia, joiden kertoimia sekä testiaineiston tuloksia vertailemalla saatiin viitteitä mallin validiteetista.

### **3.4 Tilavuuskartta ja MVMI-tuote**

Puulajikohtaisten tilavuusmallien laskemisen jälkeen niitä sovellettiin Sentinel-2-kuvan tiilen alueella (T35VLH, T35VLG, T35VMG ja T35VMH). Ensin satelliittikuvasta irrotettiin metsäalueet perustuen maankäyttöä ja maanpeitettä kuvaavaan aineistoon: Corine Maanpeite 2012, 20 m, joka on Suomen ympäristökeskuksen (Syke) tuottama (ladattavissa: [http://www.syke.fi/fi-FI/Avoin\\_tieto/Paikkatietoaineistot](http://www.syke.fi/fi-FI/Avoin_tieto/Paikkatietoaineistot)). Aineistossa metsäalueita vastaavat luokat, joiden luku on yli 20 ja alle 37.

Satelliittikuvan metsäalueet luokiteltiin Random Forest -luokittelun avulla seuraaviin luokkiin: mänty, kuusi, lehtipuu, havusekametsä ja sekametsä. Random Forest -luokittelun opeusaineistona on käytetty samaa Metsäkeskuksen koeala-aineistoa kuin tilavuusmallien muodostamisessa. Lopuksi puulajiluokkaa vastaavalla tilavuusmallilla laskettiin pikselikohtainen tilavuusarvio.

Tämän tutkimuksen tilavuusarvioita verrattiin monilähteisen valtakunnan metsien inventoinnin tilavuusarvioon. Vertailussa käytetty MVMI-tuotteen tilavuusarvio on ladattu Luonnonvarakeskuksen aineistonlatauspalvelusta (saatavissa: <http://kartta.luke.fi/open-data/valinta.html>) karttalehtien K4, L6 ja M4 alueilta. Ladatun teeman nimi oli: Tilavuus, puusto yhteensä 2015 ( $\text{m}^3/\text{ha}$ ).

Tämän tutkimuksen sekä MVMI-tuotteen tilavuusarviot luokiteltiin kuuteen (6) luokkaan, joiden rajat olivat  $0\text{--}50 \text{ m}^3/\text{ha}$ ,  $>50\text{--}100 \text{ m}^3/\text{ha}$ ,  $>100\text{--}150 \text{ m}^3/\text{ha}$ ,  $>150\text{--}200 \text{ m}^3/\text{ha}$ ,  $>200\text{--}250 \text{ m}^3/\text{ha}$  ja  $>250 \text{ m}^3/\text{ha}$ . Yli  $250 \text{ m}^3/\text{ha}$  tilavuuksia ei luokiteltu tarkemmin, sillä arviointi onnistuu parhaiten metsille, joiden tilavuus on noin  $100\text{--}300 \text{ m}^3/\text{ha}$  (Holmgren et al. 2000). Luokittelun jälkeen tilavuusarvioille suoritettiin visuaalinen tarkastelu ja laskettiin erotuskartta tämän tutkimuksen sekä MVMI-tuotteen tuloksien välillä.



## 4 Tulokset ja niiden tarkastelu

### 4.1 Pearsonin korrelaatiokertoimet

Koko koeala-aineistoon perustuen laskettiin kanavien väliset Pearsonin korrelaatiokertoimet (Taulukko 6). Korrelaatiokertoimet antavat viitteitä mahdollisesta kanavien välisestä multikollineaarisuudesta, joka vaikuttaa tilavuusmallissa käytettävien selittävien muuttujien valintaan. Sentinel-2-kanavien väliset korrelaatiokertoimet olivat kaikissa tapauksissa koko aineistolla sekä eri puulajiluokilla positiivisia. Vastaavasti Holmgren et al. (2000) raportoivat Landsat 5 TM -satelliitin kanavien välisestä positiivisesta korrelaatiosta, joka oli voimakkain kahden SWIR-aallonpituusalueen kanavan välillä. Myös tässä tutkimuksessa havaittiin SWIR-aallonpituusalueen kanavien (B11, B12) välillä voimakas korrelaatio sekä koko aineistolla että kaikilla puulajiluokilla.

Koko aineistolle lasketusta korrelaatiomatriisista oli havaittavissa kaksi keskenään korreloivaa ryhmää, joista ensimmäiseen ryhmään (taulukossa 6 punaisella taustavärillä) kuului kanavia näkyvän valon (B2, B3, B4), punaisen reunan (B5) sekä SWIR-aallonpituusalueilta (B11, B12) ja toiseen ryhmään (taulukossa 6 keltaisella taustavärillä) punaisen reunan (B6, B7) sekä lähi-infrapunaa aallonpituusalueen kanavia (B8, B8a). Ensimmäisen ryhmän kanavien korrelaatiokertoimet olivat välillä 0,8-0,96 ja toisen ryhmän 0,96-0,99.

*Taulukko 6. Korrelaatiomatriisi, jossa Sentinel-2-kanavien väliset korrelaatiokertoimet koko koeala-aineistosta laskettuna.*

	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,88	0,89	0,80	0,52	0,45	0,46	0,45	0,74	0,81
B3	0,88	1,00	0,90	0,96	0,74	0,66	0,67	0,67	0,86	0,87
B4	0,89	0,90	1,00	0,87	0,50	0,42	0,42	0,43	0,81	0,91
B5	0,80	0,96	0,87	1,00	0,79	0,72	0,71	0,73	0,91	0,89
B6	0,52	0,74	0,50	0,79	1,00	0,99	0,96	0,99	0,81	0,65
B7	0,45	0,66	0,42	0,72	0,99	1,00	0,96	1,00	0,77	0,59
B8	0,46	0,67	0,42	0,71	0,96	0,96	1,00	0,97	0,75	0,58
B8a	0,45	0,67	0,43	0,73	0,99	1,00	0,97	1,00	0,78	0,60
B11	0,74	0,86	0,81	0,91	0,81	0,77	0,75	0,78	1,00	0,95
B12	0,81	0,87	0,91	0,89	0,65	0,59	0,58	0,60	0,95	1,00

Pearsonin korrelaatiokertoimet laskettiin myös puulajiluokittain (Liite 1, Taulukot 2-6). Männyllä oli havaittavissa samankaltaiset keskenään korreloivat kanavaryhmät (Liite 1, Taulukko 2) kuin koko aineistolla. Myös kanavien väliset korrelaatiokertoimet olivat samaa suuruusluokkaa. Näkyvän valon (B2, B3, B4), ensimmäisen punaisen reunan (B5) sekä SWIR-aallonpituusalueiden (B11, B12) kanavien välillä kertoimet olivat 0,77-0,95. Ryhmän alhaisin kerroin oli sinisen valon aallonpituutta vastaavan kanavan (B2) ja punaisen reunan ensimmäisen kanavan (B5) välillä (0,77), kun taas korkein kerroin oli vihreän valon aallonpituutta vastaan kanavan (B3) ja punaisen reunan ensimmäisen kanavan (B5) välillä (0,95). Toisessa ryhmässä, jossa oli sekä punaisen reunan (B6, B7) että lähi-infrapunaa kanavia (B8, B8a), korrelaatiokertoimet olivat 0,94-0,99.

Kuusella ei ollut havaittavissa yhtä selkeitä keskenään korreloivia kanavaryhmiä (Liite 1, Taulukko 3). Kuten männyllä myös kuusella näkyvän valon kanavat (B2, B3, B4) ja punai-



sen reunan ensimmäinen kanava (B5) korreloivat voimakkaasti keskenään (0,9-0,98). Tämän lisäksi kaikki punaisen reunan (B5, B6, B7) ja lähi-infrapun kanavat (B8, B8a) korreloivat keskenään (0,88-1,00). Ryhmät eivät olleet selkeästi erottuvia, sillä niiden välillä esiintyi keskenään korreloivia kanavia. Esimerkiksi vihreän valon aallonpituuden (B3), ensimmäisen punaisen reunan (B5), ensimmäisen SWIR-aallonpituusalueen (B11) kanavat korreloivat kaikkien kanavien kanssa vähintään yli 0,85 korrelaatiokertoimella.

Lehtipuulla yli 0,9 korrelaatiokertoimet esiintyivät seuraavien kolmen eri kanavaryhmän välillä, joista ensimmäinen muodostui kahdesta näkyvän valon kanavasta (B3, B4) sekä punaisen reunan ensimmäisestä kanavasta (B5), toinen ryhmä muodostui toisesta ja kolmannesta punaisen reunan kanavasta (B6, B7) sekä kapeasta lähi-infrapun kanavasta (B8a) ja kolmas ryhmä molemmista SWIR-aallonpituuden kanavista (B11, B12) (Liite 1, Taulukko 4). Poikkeuksena oli ensimmäisen ryhmän kanavat näkyvän valon aallonpituusalueella (B3, B4), joiden välinen korrelaatiokerroin oli 0,89 kertoimen 0,9 sijaan.

Havusekametsää kuvaavista kanavista sinisen valon aallonpituusalueen kanava (B2) korreloi heikosti muiden kanavien kanssa. Kaikki kertoimet kanavan B2 ja muiden kanavien välillä jäivät alle 0,84 (Liite 1, Taulukko 5). Ensimmäinen keskenään korreloiva ryhmä kertoimilla 0,91-0,96 muodostui näkyvän valon vihreästä ja punaisesta kanavasta (B3, B4) sekä punaisen reunan ensimmäisestä kanavasta (B5). Toisen ryhmän kertoimet olivat 0,96-0,99 ja siihen kuului punaisen reunan (B6, B7) sekä lähi-infrapun kanavia (B8, B8a). Kolmas kanavaryhmä muodostui SWIR-aallonpituusalueen kanavista (B11, B12) (0,98). Tämän lisäksi punaisen aallonpituusalueen kanava (B4) ja toinen SWIR-aallonpituusalueen kanava (B12) sekä punaisen reunan ensimmäinen kanava (B5) ja ensimmäinen SWIR-aallonpituusalueen kanava (B11) korreloivat voimakkaasti keskenään (0,91-0,92).

Sekametsällä muodostui kaksi keskenään korreloivaa kanavaryhmää, joista ensimmäiseen kuului kertoimilla 0,88-0,95 näkyvän valon kanavat (B2, B3, B4), punaisen reunan ensimmäinen kanava (B5) sekä SWIR-aallonpituusalueen kanavat (B11, B12) ja toiseen kuului kertoimilla 0,94-0,99 punaisen reunan toinen (B6) ja kolmas (B7) kanava sekä lähi-infrapun kanavat (B8, B8a) (Liite 1, Taulukko 6). Sinisen valon aallonpituusalueen kanava (B2) korreloi heikosten ensimmäisessä ryhmässä punaisen reunan ensimmäisen kanavan (B5) sekä ensimmäisen SWIR-aallonpituusalueen (B11) kanavan kanssa (0,79).

Kaikilla puulajiluokilla löytyi vähintään kaksi ryhmää, joista toiseen kuului näkyvän valon aallonpituuden kanavia ja toiseen punaisen reunan sekä lähi-infrapun aallonpituuden kanavia. Punaisen reunan ensimmäinen kanava (B5) korreloi vähintään yhden näkyvän valon kanavan kanssa. SWIR-aallonpituuden kanavat (B11, B12) korreloivat vaihtelevasti muiden kanavien kanssa. Kaikilla puulajiluokilla lukuun ottamatta lehtipuuta korrelaatiokerroin oli yli 0,9 ensimmäisen SWIR-aallonpituusalueen (B11) ja punaisen reunan ensimmäisen (B5) kanavan sekä toisen SWIR- (B12) ja punaisen valon aallonpituusalueen (B4) kanavan välillä.

Kanavien välisissä korrelaatioissa oli eroja eri puulajiluokissa ja kertoimien suuruus vaihteli välillä 0,13-1. Alhaisin korrelaatiokerroin (0,13) oli puulajiluokassa mänty punaisen valon aallonpituusalueen (B4) ja punaisen reunan kolmannen (B7) kanavan välillä. Korkein korrelaatiokerroin oli puulajiluokassa kuusi punaisen reunan kolmannen (B7) ja kapean lähi-infrapun (B8a) kanavan välillä (1,00). Kuusen kanavien välillä löytyi enemmän keskenään

voimakkaasti (yli 0,9) korreloivia kanavia kuin muilla puulajiluokilla. Myös havusekametsällä keskenään korreloivia kanavia oli melko paljon. Havusekametsän noin puolella havainnoista kuusen prosenttiosuus oli välillä 50 % ja < 75 %, mikä voi osittain selittää havusekametsän samankaltaisia tuloksia kuusen kanssa.

Kanavien välisen vertailun jälkeen laskettiin tilavuuden tai tilavuuden luonnollisen logaritmin ja kanavien väliset Pearsonin korrelaatiokertoimet (Taulukko 7). Kaikki korrelaatiokertoimet olivat etumerkiltään negatiivisia, mikä tarkoittaa mitatun heijastussuhteen pienene mistä puuston tilavuuden kasvaessa. Korrelaatiokerroin tilavuuden luonnollisen logaritmin ja kanavan välillä oli suurempi kuin tilavuuden ja kanavan välillä. Tämä oli havaittavissa sekä koko aineistolla että kaikissa puulajiluokissa.

Koko aineistolla laskettuna tilavuus korreloi parhaiten punaisen reunan ensimmäisen (B5: -0,70), vihreän valon aallonpituusalueen (B3: -0,67) ja punaisen reunan toisen (B6: -0,65) kanavien kanssa, kun taas tilavuuden luonnollisen logaritmin kanssa korreloi punaisen reunan ensimmäinen (B5: -0,84), vihreän valon aallonpituusalueen (B3: -0,81) ja ensimmäinen SWIR-aallonpituusalueen (B11: -0,74) kanavat. Vaikka vihreän valon aallonpituutta kuvaava kanava (B3) ja ensimmäinen punaisen reunan kanava (B5) olivat parhaiten korreloivat kanavat, niitä ei voitu käyttää samassa regressioanalyysissä voimakkaan keskinäisen korrelaatiokertoimen vuoksi (0,96).

*Taulukko 7. Tilavuuden (V) ja tilavuuden luonnollisen logaritmin (lnV) sekä Sentinel-2-kanavien korrelaatiomatriisi. Keltaisella pohjalla kertoimet  $\leq -0,7$ , oranssilla  $\leq -0,8$  ja punaisella  $\leq -0,9$ .*

		B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
Koko ai-	Koko ai-	-0,52	-0,67	-0,54	-0,70	-0,65	-0,61	-0,63	-0,62	-0,64	-0,59
neisto V	neisto lnV	-0,61	-0,81	-0,65	-0,84	-0,71	-0,65	-0,68	-0,67	-0,74	-0,70
Mänty V	Mänty lnV	-0,42	-0,59	-0,47	-0,65	-0,53	-0,47	-0,52	-0,49	-0,56	-0,48
Kuusi V	Kuusi lnV	-0,64	-0,73	-0,60	-0,74	-0,76	-0,76	-0,75	-0,76	-0,68	-0,63
Lehtipuu V	Lehtipuu lnV	-0,35	-0,47	-0,41	-0,48	-0,17	-0,10	-0,26	-0,14	-0,41	-0,46
Havuseka-	Havuseka-	-0,45	-0,69	-0,56	-0,72	-0,80	-0,81	-0,79	-0,80	-0,58	-0,51
metsä V	metsä lnV	-0,59	-0,81	-0,67	-0,82	-0,85	-0,83	-0,83	-0,83	-0,67	-0,61
Sekametsä V	Sekametsä lnV	-0,42	-0,61	-0,49	-0,66	-0,62	-0,59	-0,58	-0,60	-0,61	-0,55
		-0,54	-0,80	-0,68	-0,85	-0,70	-0,65	-0,60	-0,66	-0,78	-0,76

Männyn tilavuuden kanssa korreloivaa kolme parasta kanavaa olivat punaisen reunan ensimmäinen kanava (B5: -0,65), vihreän valon aallonpituusalueen kanava (B3: -0,59) ja ensimmäinen SWIR-aallonpituusalueen kanava (B11: -0,57). Tilavuuden luonnollisen logaritmin kanssa kaksi parasta kanavaa olivat samat kuin tilavuudella, mutta niiden korrelaatiokertoimet olivat suuremmat (B5: -0,75 ja B3: -0,69). Kolmanneksi paras kanava oli punaisen reunan toinen kanava (B6: -0,68). Parhaiten korreloivien kanavien B5 ja B3 keskinäinen

korrelaatio oli voimakas (0,95), joten niitä ei voinut valita samaan regressiomalliin selittäviksi muuttujiksi.

Kuusen tilavuus korreloi parhaiten punaisen reunan toisen ja kolmannen (B6, B7) sekä kapean lähi-infrapun (B8a) aallonpituuden kanavien kanssa, joilla kaikilla oli sama korrelaatiokerroin (-0,76). Tilavuuden luonnollisen logaritmin kanssa parhaiten korreloi punaisen reunan ensimmäinen kanava (B5) kertoimella -0,92, joka oli paras korrelaatiokerroin kaikista puulajiluokista. Seuraavaksi parhaiten korreloivia kanavia olivat kapean lähi-infrapun kanava (B8a: -0,9) sekä vihreän valon aallonpituusalueen (B3), punaisen reunan toisen (B6) ja ensimmäinen SWIR-aallonpituusalueen (B11) kanavat, joilla kaikilla korrelaatiokerroin oli -0,89. Ainoastaan punaisen reunan toinen kanava (B6) esiintyi sekä tilavuuden että tilavuuden luonnollisen logaritmin parhaiten korreloivien kanavien kärjessä. Muutoin kuusen tilavuuden ja tilavuuden luonnollisen logaritmin välillä ei esiintynyt yhtäläisyyksiä kolmen parhaiten korreloivien kanavien välillä.

Lehtipuulla kanavien korrelaatio oli heikoin, kun verrataan sitä muiden puulajiluokkien korrelaatiokertoimiin. Tilavuus korreloi parhaiten punaisen reunan ensimmäisen (B5: -0,48), vihreän valon aallonpituusalueen (B3: -0,47) sekä toisen SWIR-aallonpituusalueen (B12: -0,46) kanavien kanssa. Tilavuuden luonnollisen logaritmin kanssa kaksi ensimmäistä parhaiten korreloivaa kanavaa olivat vastaavat kuin tilavuudella (B5 ja B3), mutta niiden korrelaatiokerroin oli suurempi (molemmilla -0,77). Kolmas parhaiten korreloiva kanava oli punaisen aallonpituusalueen kanava (B4) kertoimella -0,69.

Havusekametsän tilavuus korreloi punaisen reunan toisen ja kolmannen (B6, B7) sekä lähi-infrapun aallonpituusalueen kanavien (B8, B8a) kanssa vaihtelevassa järjestyksessä sekä tilavuudella että tilavuuden luonnollisella logaritmillä. Tilavuuden kanssa korreloivat kanavat suuruusjärjestyksessä olivat B7 (-0,81), B6 ja B8a (-0,80) sekä B8 (-0,79). Tilavuuden luonnollisella logaritmillä parhaiten korreloiva kanava oli B6 (-0,85), seuraavat kolme kanavaa B7, B8 ja B8a korreloivat samalla kertoimella -0,83. Muista puulajiluokista poiketen parhaiten korreloivien kanavien joukossa ei esiintynyt punaisen reunan ensimmäistä kanavaa (B5).

Sekametsän tilavuuden sekä tilavuuden luonnollisen logaritmin kanssa parhaiten korreloiva kanava oli punaisen reunan ensimmäinen kanava B5 (tilavuus: -0,66 ja tilavuuden luonnollinen logaritmi: -0,85). Seuraavaksi parhaimmat tilavuuden kanssa korreloivat kanavat olivat punaisen reunan toinen (B6: -0,62) ja vihreän valon aallonpituusalueen (B3: -0,61) kanavat. Tilavuuden luonnollisen logaritmin kanssa korreloivat seuraavat kanavat olivat vuorostaan vihreän valon aallonpituusalueen (B3: -0,80) sekä ensimmäisen SWIR-aallonpituusalueen (B11: -0,78) kanavat.

Heikoin korrelaatiokerroin oli lehtipuun tilavuuden ja punaisen reunan kolmannen kanavan (B7) välillä (-0,10). Vastaavasti tilavuuden luonnollisen logaritmin kanssa heikoiten korreloi punaisen reunan kolmas kanava (B7) puulajiluokassa lehtipuu (-0,16). Tämän perusteella voidaan todeta, että punaisen reunan kolmas kanava ei sovellu lehtipuun tilavuuden arvioimiseen.

Tilavuuden kanssa voimakkaimmin korreloi punaisen reunan kolmas kanava (B7) puulajiluokassa havusekametsä (-0,81). Suurin korrelaatiokerroin oli kuusen tilavuuden luonnolli-

sen logaritmin ja punaisen reunan ensimmäinen kanavan (B5) välillä (-0,92). Pearsonin korrelaatiokertoimia tarkastellessa useasti toistuvia parhaimpia kanavia olivat vihreän valon aallonpituusalueen kanava (B3) sekä punaisen reunan kanavat (B5, B6, B7), joista erityisesti punaisen reunan ensimmäinen kanava (B5). Kuitenkin kanavien sekä tilavuuden välistä korrelaatioita on tarkasteltava puulajiluokittain, sillä esimerkiksi punaisen reunan kolmas kanava (B7) korreloi sekä heikosti että voimakkaasti tilavuuden kanssa riippuen puulajiluokasta. Tämä havainto tukee tilavuusarvioinnin toteuttamista puulajikohtaisesti.

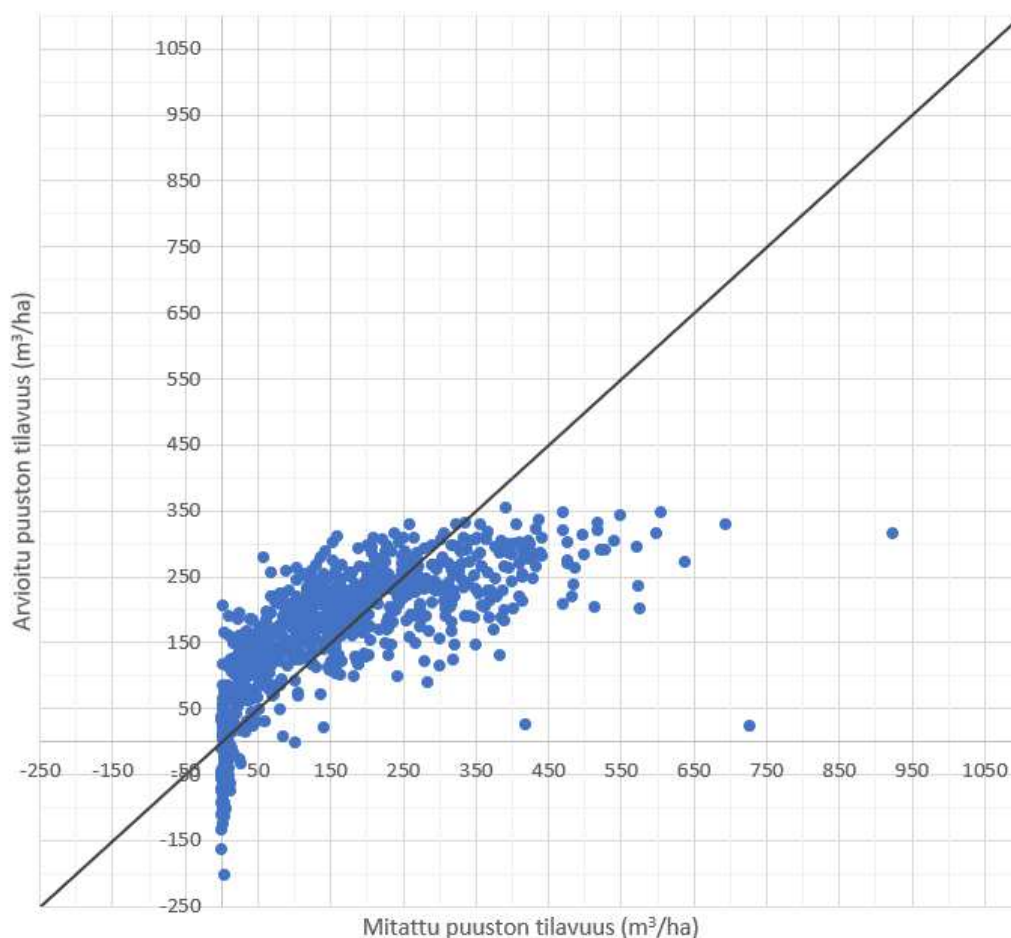
Chrysafis et al. (2017) myös havaitsivat myös negatiivisen korrelaation Sentinel-2-kanavien ja Välimeren puuston tilavuuden kanssa. Heidän tutkimuksessaan alhaisin korrelaatiokerroin oli punaisen valon aallonpituusalueen kanavan (B4) sekä tilavuuden välillä (-0,04) ja korkein kerroin ensimmäisen SWIR-aallonpituusalueen kanavan (B11) kanssa (-0,61). Ero tämän tutkimuksen sekä Chrysafiksen et al. (2017) tutkimuksen parhaimpien kanavien välillä voi osittain selittyä erilaisilla puulajeilla sekä kasvillisuusvyöhykkeellä. Chrysafiksen et al. (2017) tutkimuskohde sijoittui homogeeniseen metsäalueeseen Kreikassa, kun taas tämän tutkimuksen kohde kuului pohjoiseen havumetsävyöhykkeeseen.

## **4.2 Regressioanalyysin tulokset**

Pearsonin korrelaatiokertoimiin perustuen laskettiin regressiomallit parhaiten korreloiville kanaville sekä niistä muodostetuille kanavapareille. Regressioanalyysiin valittujen kanavaparien välinen korrelaatiokerroin ei saanut ylittää 0,9, jotta selittävien muuttujien välillä ei esiintynyt multikollineaarisuutta. Liitteen 2 taulukossa 1 on regressioanalyysin tulokset, kun selitettävänä muuttujana oli tilavuus. Liitteen 2 taulukossa 2 regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi.

### **4.2.1 Selitettävänä muuttujana tilavuus**

Koko aineiston tilavuuden arvioinnissa kahden muuttujan regressiomalli käyttäen punaisen reunan ensimmäistä (B5) sekä toista (B6) kanavaa tuotti suurimman korjatun  $R^2$ -luvun (0,51) sekä pienimmät arvot seuraaville tunnusluvuille RMSE (97,4 m<sup>3</sup>/ha), nRMSE (60,0 %), MdAE (60,2 m<sup>3</sup>/ha), nMdAE (44,1 %) (Kuva 11). Yhden muuttujan lineaarisessa regressiomallissa paras yksittäinen kanava oli punaisen reunan ensimmäinen kanava (B5), kun tarkasteltiin korjattua  $R^2$ -lukua sekä keskineliövirheen neliöjuuren että normalisoidun keskineliövirheen neliöjuuren arvoa. Absoluuttinen mediaanivirhe oli pienin punaisen reunan toisen kanavan (B6) regressiomallissa. Mallin arvioima maksimiarvo jäi kaikilla koko aineistolle lasketuilla malleilla alle 363,0 m<sup>3</sup>/ha, vaikka koko aineiston havaittu maksimitilavuus oli 924,0 m<sup>3</sup>/ha. Lisäksi kaikilla malleilla minimiarvot olivat negatiivisia alhaisimmillaan -246,43 m<sup>3</sup>/ha.



Kuva 11. Koko aineiston regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,51, RMSE 97,4 m<sup>3</sup>/ha, nRMSE 60,0 %, MdAE 60,2 m<sup>3</sup>/ha, nMdAE 44,1 %, maksimitilavuus 354,4 m<sup>3</sup>/ha sekä minimi-tilavuus -202,9 m<sup>3</sup>/ha.

Kuten koko aineistolla myös männyn tilavuuden arvioinnissa paras yksittäinen kanava oli punaisen reunan ensimmäinen kanava (B5) ja kahden muuttujan mallissa parhaimmat kanavat olivat punaisen reunan ensimmäinen ja toinen kanava (B5, B6). Mallia kuvaavien tunnuslukujen välillä ei ollut suurta eroa. Kanavalla B5 muodostetulla mallilla korjattu  $R^2$  -luku oli 0,42, RMSE 70,6 m<sup>3</sup>/ha, nRMSE 46,6%, MdAE 44,7 m<sup>3</sup>/ha ja nMdAE 30,5 %. Vastaavat luvut kanavien B5 ja B6 avulla muodostetulla mallilla olivat 0,452, 68,45 m<sup>3</sup>/ha, 45,1 %, 44,4 m<sup>3</sup>/ha ja 30,3 %. Myös männyn tilavuutta arvioidessa mallien maksimi- ja minimiarvot jäivät alle 288,3 m<sup>3</sup>/ha. Lisäksi minimiarvot olivat negatiivisia ja alhaisimmillaan -85,6 m<sup>3</sup>/ha.

Paras yksittäinen kanava kuusen tilavuuden arvioimiseksi oli punaisen reunan toinen kanava (B6). Kahden selittävän muuttujan mallissa kanavat olivat vihreän valon aallonpituusalueen (B3) ja punaisen reunan kolmas (B7) kanava. Vastaavasti kuin männyllä myös kuusella parhaimman yhden selittävän muuttujan mallin että kahden selittävän muuttujan mallin välillä ei ollut suurta eroa. Kanavan B6 muodostaman mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,58, RMSE 109,7 m<sup>3</sup>/ha, nRMSE 49,9 %, MdAE 71,1 m<sup>3</sup>/ha, nMdAE 35,5 %, maksimitilavuus 417,0 m<sup>3</sup>/ha ja minimi-tilavuus -137,4 m<sup>3</sup>/ha. Vastaavat luvut kanavaparin B3 ja B7 mallissa olivat korjattu  $R^2$  -luku 0,59, RMSE 108,6 m<sup>3</sup>/ha, nRMSE 49,4 %, MdAE 70,5 m<sup>3</sup>/ha, nMdAE 35,2 %, maksimitilavuus 413,0 m<sup>3</sup>/ha ja minimi-tilavuus -117,7 m<sup>3</sup>/ha.

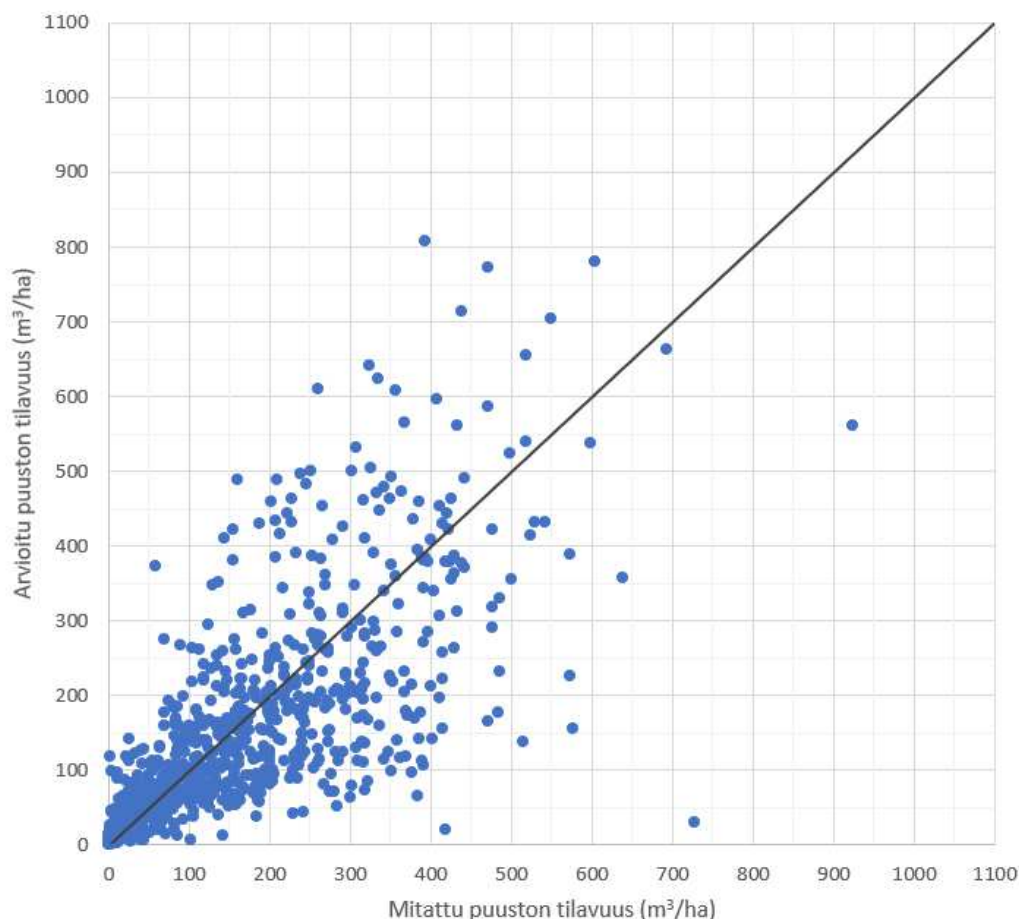
Lehtipuun tilavuuden arvioimiseen soveltuvia kahden kanavan malleja muodostui yksi, jossa selittävinä muuttujina olivat vihreän valon aallonpituusalueen (B3) sekä toisen SWIR-aallonpituusalueen kanavat (B12). Muilla kanavapareilla regressiokertoimien etumerkki vaihtui negatiivisesta positiiviseen, mikä viittaa epävakaiseen malliin. Kanavaparin B3 ja B12 selitysaste jäi kaikista tilavuusmalleista heikoimmaksi korjatun  $R^2$ -luvun ollessa 0,22. Muut tunnusluvut olivat RMSE 93,3 m<sup>3</sup>/ha, nRMSE 98,6 % MdAE 45,4 m<sup>3</sup>/ha ja nMdAE 66,9 %. Yhden kanavan mallit eivät poikenneet merkittävästi kahden kanavan mallista. Korjatun  $R^2$ -luvun sekä keskineliövirheen neliöjuuren perustella paras yksittäinen malli syntyi kanavalla B5, jonka tunnusluvut olivat korjattu  $R^2$ -luku 0,23 RMSE 93,0 m<sup>3</sup>/ha, nRMSE 98,8 % MdAE 44,6 m<sup>3</sup>/ha ja MdAE 65,6 %. Mallien maksimitilavuudet jäivät alle 184,7 m<sup>3</sup>/ha ja minimi-tilavuudet olivat alimmillaan -95,0 m<sup>3</sup>/ha.

Havusekametsän tilavuuden arvioimiseen luotettavasti ei löytynyt kahden kanavan mallia, joka olisi täyttänyt regressiomallille asetetut kriteerit. Kahden kanavan malleissa toinen selittävästä muuttujista oli joko tilastollisesti merkitsemätön tai mallin regressiomuuttujien etumerkit viittasivat positiivisilla etumerkeillään epävakaiseen malliin. Parhaimman yhden kanavan mallin selittävä muuttuja oli punaisen reunan kolmas kanava (B7). Kanavan B7 tilavuusmallin korjattu  $R^2$ -luku oli 0,66, RMSE 78,8 m<sup>3</sup>/ha, nRMSE 41,3 %, MdAE 54,9 m<sup>3</sup>/ha, nMdAE 31,4 %, maksimitilavuus 362,0 m<sup>3</sup>/ha ja minimi-tilavuus -91,5 m<sup>3</sup>/ha.

Sekametsän tilavuuden arvioinnissa selitysasteen ja keskineliövirheen neliöjuuren perusteella paras yksittäinen kanava oli punaisen reunan ensimmäinen kanava (B5) (korjattu  $R^2$ -luku 0,44, RMSE 104,6 m<sup>3</sup>/ha, nRMSE 70,6 %) ja paras kanavapari oli punaisen reunan ensimmäinen ja toinen kanava (B5, B6) (korjattu  $R^2$ -luku 0,47, RMSE 101,5 m<sup>3</sup>/ha, nRMSE 68,5 %). Absoluuttinen mediaanivirhe oli pienin yhden muuttujan mallilla B6 (MdAE 68,5 m<sup>3</sup>/ha ja nMdAE 62,6 %) ja kahden muuttujan mallilla B3 ja B6 (MdAE 64,6 m<sup>3</sup>/ha ja nMdAE 59,1 %). Näillä malleilla arvioidut maksimi- ja minimiarvot olivat välillä 279,4 ja 337,5 m<sup>3</sup>/ha sekä minimi- ja maksimi-tilavuudet olivat välillä -56,0 ja -215,7 m<sup>3</sup>/ha.

#### 4.2.2 Selitettävänä muuttujana tilavuuden luonnollinen logaritmi

Koko aineistolla muodostetussa regressiomallissa, jossa selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi, parhaimmat kanavat olivat vastaavat, kun selitettävänä muuttujana oli tilavuus. Erona oli tunnusluvut, joista korjattu  $R^2$ -luku punaisen reunan ensimmäisellä kanavalla (B5) oli 0,71 sekä punaisen reunan ensimmäisen ja toisen kanavan (B5, B6) mallilla 0,72. Muut tunnusluvut kanavan B5 mallille olivat RMSE 104,0 m<sup>3</sup>/ha, nRMSE 64,0 %, MdAE 39,3 m<sup>3</sup>/ha, nMdAE 28,8 %, maksimitilavuus 715,9 m<sup>3</sup>/ha sekä minimi-tilavuus 0,6 m<sup>3</sup>/ha. Kahden kanavan malli oli tunnusluvuiltaan parempi, mutta erot mallien välillä eivät olleet suuria, esimerkiksi nRMSE- ja nMdAE-arvojen ero mallien välillä oli noin 1 prosenttiyksikköä. Kanavien B5 ja B6 mallin tunnusluvut olivat RMSE 102,2 m<sup>3</sup>/ha, nRMSE 63,0 %, MdAE 38,2 m<sup>3</sup>/ha, nMdAE 28,0 %, maksimitilavuus 807,1 m<sup>3</sup>/ha sekä minimi-tilavuus 0,7 m<sup>3</sup>/ha (Kuva 12).



Kuva 12. Koko aineiston regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,72, RMSE 102,2 m<sup>3</sup>/ha, nRMSE 63,0 %, MdAE 38,2 m<sup>3</sup>/ha, nMdAE 28,0 %, maksimitilavuus 807,1 m<sup>3</sup>/ha sekä minimi-tilavuus 0,7 m<sup>3</sup>/ha.

Männyn tilavuuden selittämiseen soveltui parhaiten punaisen reunan ensimmäinen kanava (B5) tai kahden kanavan mallissa punaisen reunan ensimmäinen ja toinen kanava (B5, B6). Kanavalla B5 muodostetun mallin korjattu  $R^2$  -luku oli 0,55, kun taas kanavaparin B5 ja B6 mallissa korjattu  $R^2$  -luku oli 0,67. Kanavan B5 mallissa muut tunnusluvut osoittivat mallin soveltuvan paremmin männyn tilavuuden arvioimiseen (RMSE 75,6 m<sup>3</sup>/ha, nRMSE 49,8 %, MdAE 37,7 m<sup>3</sup>/ha, nMdAE 25,7 %, maksimitilavuus 417,9 m<sup>3</sup>/ha ja minimi-tilavuus 5,5 m<sup>3</sup>/ha). Kanavien B5 ja B6 mallissa vastaavat tunnusluvut olivat RMSE 85,7 m<sup>3</sup>/ha, nRMSE 56,5 %, edAE 40,0 m<sup>3</sup>/ha, nMdAE 27,3 %, maksimitilavuus 709,0 m<sup>3</sup>/ha ja minimi-tilavuus 4,5 m<sup>3</sup>/ha.

Kuusen tilavuuden arvioinnissa paras yksittäinen kanava oli myös punaisen reunan ensimmäinen kanava (B5). Kahden kanavan mallissa keskineliövirheen neliöjuuren ja absoluuttisen mediaanivirheen mukaan paras malli syntyi vihreän valon aallonpituusalueen (B3) sekä punaisen reunan kolmannella kanavalla (B7). Kanavalla B5 mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,84, RMSE 103,8 m<sup>3</sup>/ha, nRMSE 47,2 %, MdAE 52,2 m<sup>3</sup>/ha, nMdAE 26,0 %, maksimitilavuus 886,4 m<sup>3</sup>/ha ja minimi-tilavuus 0,6 m<sup>3</sup>/ha sekä vastaavat luvut kanavien B3 ja B7 mallille olivat korjattu  $R^2$  -luku 0,81, RMSE 115,1 m<sup>3</sup>/ha, nRMSE 52,4 %, MdAE 45,2 m<sup>3</sup>/ha, nMdAE 22,6 %, maksimitilavuus 808,8 m<sup>3</sup>/ha ja minimi-tilavuus 1,6 m<sup>3</sup>/ha.

Lehtipuun tilavuuden arvioinnissa yhden kanavan mallin keskineliövirheen neliöjuuren sekä absoluuttisen mediaanivirheen mukaan paras kanava oli punaisen reunan ensimmäinen kanava (B5), jonka tunnusluvut olivat korjattu  $R^2$  -luku 0,59, RMSE 97,7 m<sup>3</sup>/ha, nRMSE 103,2 % MdAE 20,3 m<sup>3</sup>/ha ja nMdAE 29,8 %, maksimitilavuus 260,5 m<sup>3</sup>/ha ja minimitilavuus 0,7 m<sup>3</sup>/ha. Tutkituista kanavapareista regressioanalyysin kriteerit täytti yksi kanavapari, jossa mallin selittävinä muuttujina olivat punaisen valon aallonpituusalueen (B4) ja toisen SWIR-aallonpituusalueen (B12) kanavat. Muiden kahden kanavan mallien ongelmana esiintyi multikollinearisuutta tai toinen selittävästä muuttujista ei ollut tilastollisesti merkittävä. Kanavaparin B4 ja B12 mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,49, RMSE 100,8 m<sup>3</sup>/ha, nRMSE 106,5 %, MdAE 28,8 m<sup>3</sup>/ha ja nMdAE 42,5 %, maksimitilavuus 197,2 m<sup>3</sup>/ha ja minimitilavuus 1,0 m<sup>3</sup>/ha.

Havusekametsän tilavuuden arvioinnissa sekä kahden että yhden kanavan malleja oli useita mahdollisia. Yhden kanavan malleissa selityksasteen mukaan paras malli syntyi punaisen reunan toisella kanavalla (B6) (korjattu  $R^2$  -luku 0,72). Keskineliövirheen neliöjuuren perusteella paras kanava oli punaisen reunan kolmas kanava (B7) (RMSE 93,61 m<sup>3</sup>/ha ja nRMSE 49,1 %). Absoluuttista mediaanivirhettä tarkastelemalla paras yhden muuttujan malli syntyi kapean lähi-infrapun kanavalla (B8a) (MdAE 36,3 m<sup>3</sup>/ha ja nMdAE 20,7 %). Kahden kanavan malleista keskineliövirheen neliöjuuren ja absoluuttisen mediaanivirheen mukaan paras kanavapari oli vihreän valon aallonpituusalueen (B3) ja punaisen reunan toinen (B6) kanava. Tunnusluvut olivat korjattu  $R^2$  -luku 0,73, RMSE 93,5 m<sup>3</sup>/ha, nRMSE 49,0 % MdAE 30,2 m<sup>3</sup>/ha ja nMdAE 17,2 %. Mallien arvioimat maksimitilavuudet vaihtelivat 628,9 ja 963,5 m<sup>3</sup>/ha välillä ja minimitilavuudet vuorostaan 3,0 ja 6,3 m<sup>3</sup>/ha välillä.

Viidennen puulajiluokan, sekametsän, tilavuuden arvioimiseksi soveltui punaisen reunan ensimmäisen kanavan (B5) avulla muodostettu malli. Kanavan B5 mallin selityksaste oli suurin (korjattu  $R^2$  -luku 0,72) sekä absoluuttinen mediaanivirhe oli alhaisin (MdAE 38,1 m<sup>3</sup>/ha ja nMdAE 34,8 %). Paras keskineliövirheen neliöjuuren arvo saatiin vihreän valon aallonpituusalueen kanavan (B3) mallilla (RMSE 114,0 m<sup>3</sup>/ha, nRMSE 76,9 %). Eroa keskineliövirheen neliöjuuren arvoilla kanavien B5 ja B3 mallien välillä oli 0,7 prosenttiyksikköä. Kanavien B5 ja B3 mallien maksimitilavuudet olivat 552,9 m<sup>3</sup>/ha ja 560,8 m<sup>3</sup>/ha sekä minimitilavuus oli molemmilla sama 0,4 m<sup>3</sup>/ha. Kahden kanavan mallissa sekametsän tilavuuden arviointiin soveltui vihreän valon aallonpituusalueen kanava (B3) ja punaisen reunan toinen kanava (B6). Muut kanavaparit eivät täyttäneet kriteeriä, jossa molempien kanavien tulee olla tilastollisesti merkittäviä. Kanavaparin B3 ja B6 mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,65 RMSE 117,6 m<sup>3</sup>/ha, nRMSE 79,4 % MdAE 36,2 m<sup>3</sup>/ha, nMdAE 33,1 %, maksimitilavuus 724,4 m<sup>3</sup>/ha ja minimitilavuus 0,3 m<sup>3</sup>/ha.

### **4.3 Regressioanalyysin tulosten tarkastelu ja mallin valinta**

Paras regressiomalli männyn, kuusen, lehtipuun sekä sekametsän tilavuuden arvioimiseksi muodostui punaisen reunan ensimmäisellä kanavalla (B5). Koko aineiston tapauksessa paras malli muodostui punaisen reunan ensimmäisellä sekä toisella kanavalla (B5, B6). Havusekametsän tilavuuden arvioimiseen paras kanavapari oli vihreän valon aallonpituutta kuvaava kanava (B3) sekä punaisen reunan toinen kanava (B6). Punaisen reunan ensimmäinen kanava korostui myös Korhosen et al. (2017) tutkimuksessa, jossa he arvioivat metsän latvuspeittoa sekä lehtialaindeksiä. Lehtialaindeksin sekä runkotilavuuden välillä on voimakas korrelaatio (Korhonen et al. 2017), joten vaikka tutkittavat muuttujat eivät olleet samoja kuin tässä tutkimuksessa, niin voimakas korrelaatio punaisen reunan kanavien sekä lehtialaindeksin välillä antaa viitteitä mahdollisesta korrelaatiosta myös tilavuuden sekä punaisen reunan



kanavien välillä. Chrysafis et al. (2017) myös raportoi punaisen reunan ensimmäisen kanavan potentiaalista puuston tilavuusarvioinnissa. Molempien tutkimuksien tulokset viittaavat Sentinel-2 satelliitin punaisen reunan kanavien hyötyihin puuston arvioinnissa, mikä tukee tämän tutkimuksen tuloksia.

Punaisen reunan kanavien, erityisesti lähempänä punaista valoa, hyötyä voidaan selittää klorofyllien avulla. Punaisen reunan kanavat jäävät punaisen sekä lähi-infrapun aallonpituusalueen väliin. Kun biomassan määrä ja tämän seurauksena klorofyllipitoisuus on korkea, lähi-infrapun kanavat saturoituvat (Frampton et al. 2013). Saturoitumisen seurauksena suurien tilavuuksien arvioiminen lähi-infrapun kanavilla ei onnistu. Punaisen reunan kanavat vuorostaan parantavat estimaatteja, sillä biomassan kasvaessa niiden saturaatio on matalampi (Chrysafis et al. 2017) ja näin ollen tilavuuden arvioiminen paranee myös suuremmilla tilavuuksilla.

Regressiomalleissa, joissa selitettävänä muuttujana oli tilavuus, parhaimpien yhden ja kahden selittävän muuttujan mallien nRMSE-arvossa ei ollut suurta eroa (alle 2,2 prosenttiyksikköä). Kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi, erot parhaimpien yhden ja kahden selittävän muuttujan mallien nRMSE-arvojen välillä olivat suurempia puulajiluokissa mänty, kuusi sekä lehtipuu (yli 3,0 prosenttiyksikköä). Tilavuuden luonnollisen logaritmin kanssa parhaiten korreloivat kanavat korreloivat myös voimakkaasti keskenään. Tämän seurauksena kahden kanavan mallissa toinen valituista kanavista ei korreloinut yhtä hyvin tilavuuden kanssa, jolloin regressiomallin arviointikyky heikentyi verrattuna yhden kanavan regressiomalliin.

Havusekametsällä sekä koko aineistolla tilavuuden luonnollisen logaritmin parhaimman kahden kanavan mallin nRMSE oli pienempi kuin yhden kanavan mallilla. Ero ei kuitenkaan ollut merkittävä (alle 1 prosenttiyksikköä). Muilla puulajiluokilla mänty, kuusi, lehtipuu ja sekametsä parhaimman yhden kanavan mallin nRMSE oli pienempi kahden kanavan malliin verrattuna. Häme et al. (2013) myös havaitsivat trooppisen metsän biomassaa arvioivassa tutkimuksessa, että useampi optisen satelliitin kanava ei parantanut regressiomallin tuottamia tuloksia. Biomassa korreloi voimakkaasti puuston tilavuuden kanssa (Häme et al. 2013), joten tämän sekä Hämeen et al. (2013) tutkimuksen tulokset antavat viitteitä, että puuston tilavuuden tai biomassan arvioimiseen regressioanalyysillä selittäviksi muuttujiksi riittää yksi tai korkeintaan kaksi optisen satelliitin kanavaa.

Pelkästään keskineliöjuuren neliövirhettä tarkastelemalla ei tässä tutkimuksessa saatu luotettavaa kuvaa mallin toimivuudesta tilavuuden arvioinnissa. Tilavuusmallin sekä tilavuuden luonnollisen logaritmin mallin erot näkyivät selkeästi arvioidun ja mitatun tilavuuden kuvaajista (Liite 3). Ensinnäkin tilavuusmalli ilman luonnollista logaritmia tuotti negatiivisia tilavuusarvioita, jotka ovat käytännössä mahdottomia arvioita puuston tilavuudesta. Toiseksi kuvaajia tarkastelemalla oli myös havaittavissa tilavuusmallin aliarvioivan suuria tilavuuksia. Siitä huolimatta malleilla, joiden selitettävänä muuttujana oli tilavuus, RMSE oli pienempi kuin tilavuuden luonnollisen logaritmin malleilla. Poikkeuksena oli kuusi, jolla RMSE oli alhaisin tilavuuden luonnollisen logaritmin mallilla. Keskineliövirheen neliöjuuresta poiketen absoluuttinen mediaanivirhe oli pienempi kaikissa luonnollisen logaritmin mallien tapauksissa. Suurempien tilavuuksien arvioimisessa esiintyy enemmän epävarmuutta, jolloin suuremmat virhearviot kasvattavat RMSE-arvoa, kun taas MdAE kuvaa aineiston virhejakautumaa tyypillistä arvoa ja näin ollen on pienempi.

Tämän tutkimuksen tulos puoltaa luonnollisen logaritmin käyttöä tilavuuden arvioinnissa. Tilavuuden luonnollisen logaritmin ja kanavan välillä oli voimakkaampi lineaarinen suhde kuin pelkän tilavuuden ja kanavan välillä. Saman havainnon saivat myös Halme ja Tomppo (2001) sekä Trotter et al. (1997) tutkimuksessaan.

Tässä tutkimuksessa koko aineistolla muodostetun mallin nRMSE oli 63,0 %. Yksittäisten puulajien tuloksien tarkkuus vaihteli merkittävästi puulajiluokittain. Havupuuvaltaisissa metsissä tilavuusarviot olivat tarkempia kuin lehtipuuvaltaisissa metsissä. Puulajiluokittaisen mallien nRMSE oli männyllä 49,8 %, kuusella 47,2 %, lehtipuulla 103,2 %, havusekametsällä 49,0 % ja sekametsällä 77,6 %. Vastaavan eron Mäkelä ja Pekkarinen (2004) sekä Huiyan et al. (2006) havaitsivat havu- ja lehtipuiden välillä. Mäkelä ja Pekkarinen (2004) raportoivat nRMSE-arvoiksi männylle 117 %, kuuselle 81 % sekä lehtipuulle 174 % ja Huiyan et al. (2006) lehtikuuselle 51,7 %, Korean männylle 71,7 % sekä lehtipuulle 88,19 %. Mäkelä ja Pekkarinen (2004) sekä Huiyan et al. (2006) puulajikohtaiset arviot olivat jonkin verran epätarkempia kuin tässä tutkimuksessa saadut puulajikohtaiset tilavuusarviot.

Tilavuusarvio osoittautui epätarkimmaksi lehtipuulla, jonka nRMSE oli muista puulajeista poiketen yli 100 %. Myös Mäkelän ja Pekkarisen (2004) sekä Huiyan et al. (2006) tutkimuksessa lehtipuun tilavuusarvio oli kaikkein epätarkinta. Tässä tutkimuksessa havaittiin, että lehtipuun suurilla tilavuuksilla sekä taimikoilla oli molemmilla korkea heijastussuhde. Lisäksi mallin muodostamisessa oli suhteessa enemmän pienen tilavuuden taimikkoa verrattuna suuren tilavuuden lehtimetsään, jonka seurauksena malli aliarvioi suuria tilavuuksia ja nRMSE-arvo oli suuri.

Tämän tutkimuksen tulos vastaa luvussa 2.4.1 mainittujen tutkimuksien tuloksien tarkkuutta. Validointiin valittujen lopullisten mallien RMSE-arvot vaihtelivat 75,6-115,0 m<sup>3</sup>/ha, nRMSE 47,2-103,2% sekä korjattu R<sup>2</sup> -luku vaihteli 0,56-0,84 välillä. Esimerkiksi tämän tutkimuksen RMSE-arvoa vastaavia tuloksia saivat Chrysafis et al. (2017) (63,11 m<sup>3</sup>/ha) menetelmänään Random Forest -regressio, Hall et al. (2006) (110,8 m<sup>3</sup>/ha) menetelmänään BioSTRUCT, Hyyppä et al. (2000) (87,5 m<sup>3</sup>/ha) menetelmänään neuroverkko- ja regressioanalyysi sekä Mohammadi et al. (2010) (97,4 m<sup>3</sup>/ha) menetelmänään regressioanalyysi. Luetellut tutkimukset eivät ole täysin vertailukelpoisia tämän tutkimuksen kanssa, sillä muun muassa menetelmä, käytetyt aineistot sekä tutkittavan kohteen sijainti ja kasvillisuusvyöhyke vaihtelivat tutkimuksien välillä. Näissä tutkimuksissa oli käytetty Sentinel-2 (Chrysafis et al. 2017), Landsat ETM+ (Hall et al. 2006, Hyyppä et al. 2000) ja Landsat TM (Mohammadi et al. 2010) -satelliittien aineistoa, joten eroista huolimatta tämän perusteella voitaneen todeta, että Sentinel-2-satelliittiaineisto soveltuu yhtä hyvin puuston tilavuuden arvioimiseen kuin Landsat-satelliittiaineisto. Jatkotutkimuksessa voitaisiin tutkia, onko eri menetelmällä kuten kNN-menetelmällä tai Random Forest -regressiolla mahdollista parantaa Sentinel-2-satelliittiin perustuvaa tilavuusarvioinnin tarkkuutta.

Mielenkiintoista olisi myös selvittää, voiko tilavuusarvioita parantaa aikasarjaan perustavalla mallilla, sillä kasvillisuuden heijastussuhde muuttuu kasvukauden eri ajankohtina kuten metsän heijastusspektriä koskevassa luvussa 2.1 todettiin. Tämän tutkimuksen tilavuusmalli soveltuu Etelä-Suomen alueelle ja elokuussa otetuille Sentinel-2-satelliittikuvilla. Ilman lisätutkimusta mallin ei voida olettaa toimivan luottavasti esimerkiksi alkukesän tai Pohjois-Suomen satelliittikuvilla.

Luvuissa 4.2.1 sekä 4.2.2 esitellyistä malleista valittiin tutkimuksen seuraavaan vaiheeseen parhaimmat puulajiluokkaa kuvaavat mallit. Tunnuksien lisäksi mallin valinnassa huomioitiin liitteessä 3 esitetyt kuvaajat. Kuvaajat havainnollistivat, kuinka hyvin arvioitu tilavuus vastaa havaittua tilavuutta. Kaikki validointiin valitut mallit olivat muodostettu tilavuuden luonnollisen logaritmin avulla. Koko aineiston paras regressiomalli muodostui punaisen reunan ensimmäisestä (B5) sekä toisesta (B6) kanavasta. Männyn, kuusen, lehtipuun sekä sekametsän tilavuuden arvioinnissa paras malli muodostui punaisen reunan ensimmäisen kanavan (B5) avulla. Havusekametsän tilavuuden arvioinnissa parhaimman mallin tuotti vihreän valon aallonpituusalueen (B3) sekä punaisen reunan toinen (B6) kanava.

#### **4.4 Validoinnin tulokset ja niiden tarkastelu**

Koko aineiston ja puulajiluokittain lasketuille malleille suoritettiin tupla-ristiinvalidointi (Liite 4, Taulukot 1-6). Liitteen 4 taulukoissa esiintyvä opetusaineisto 1 vastasi kaikkia puulajiluokan havaintoja. Opetusaineistossa 2 oli noin puolet havainnoista ja sitä käytettiin ensin mallin opetusaineistona ja sen testiaineistona käytettiin opetusaineistoa 3. Opetusaineisto 3 muodostui opetusaineistosta 2 jäljelle jääneistä havainnoista. Taulukoissa ilmoitettu korjattu  $R^2$  -luku laskettiin mallin muodostamisen yhteydessä. Muut taulukossa esiintyvät tunnusluvut RMSE, nRMSE, MdAE, nMdAE, max V (mallin tuottama maksimitilavuus) sekä min V (mallin tuottama minimi-tilavuus) olivat tuloksia opetusaineistolla muodostetun mallin toimivuudesta testiaineistolla.

Koko aineistolla ja kaikilla puulajiluokilla eri opetusaineistojen 1-3 normalisoidun keskineliövirheen neliöjuuren (nRMSE) arvot olivat yleisesti suurempia kuin normalisoidun absoluuttisen mediaanivirheen (nMdAE). Lisäksi nMdAE:n arvojen suuruus ei vaihdellut opetusaineistojen välillä yhtä paljon kuin nRMSE:n arvot. Opetusaineistot 2 ja 3 ja niiden testiaineistot tuottivat validoinnissa sekä nRMSE:n että nMdAE:n maksimi- ja minimiarvot, jonka seurauksena opetusaineistolla 1 lasketut tunnusluvut olivat niiden väliltä. Opetusaineisto 1 muodostui kaikista havainnoista ja näin ollen on luonnollista, että opetusaineistoa vastaavan testiaineiston tuottamat tunnusluvut eivät olleet suurimpia tai pienimpiä. Koska aineisto jaettiin satunnaisesti opetusaineistoihin, on mahdollista, että toinen opetusaineisto sisälsi enemmän suurempia tai pienempiä tilavuuksia kuin toinen opetusaineisto. Tämä voi osittain selittää, miksi regressiomalleissa esiintyvät kertoimet myös vaihtelivat enemmän.

Männyn ja kuusen nRMSE:n maksimi- ja minimiarvot olivat samaa suuruusluokkaa. Suurin ero oli minimi nRMSE:n arvojen välillä (5,9 prosenttiyksikköä). Molemmissa puulajiluokissa (mänty ja kuusi) maksimi nRMSE jäi alle 55 %, joka oli muihin puulajiluokkiin verrattuna selkeästi pienempi. Lehtipuulla, havusekametsällä ja sekametsällä pienin nRMSE oli yli 66%. Lehtipuulla ja havusekametsällä maksimi nRMSE oli yli 100 %, kun taas sekametsän maksimi nRMSE jäi hieman alle 100 %.

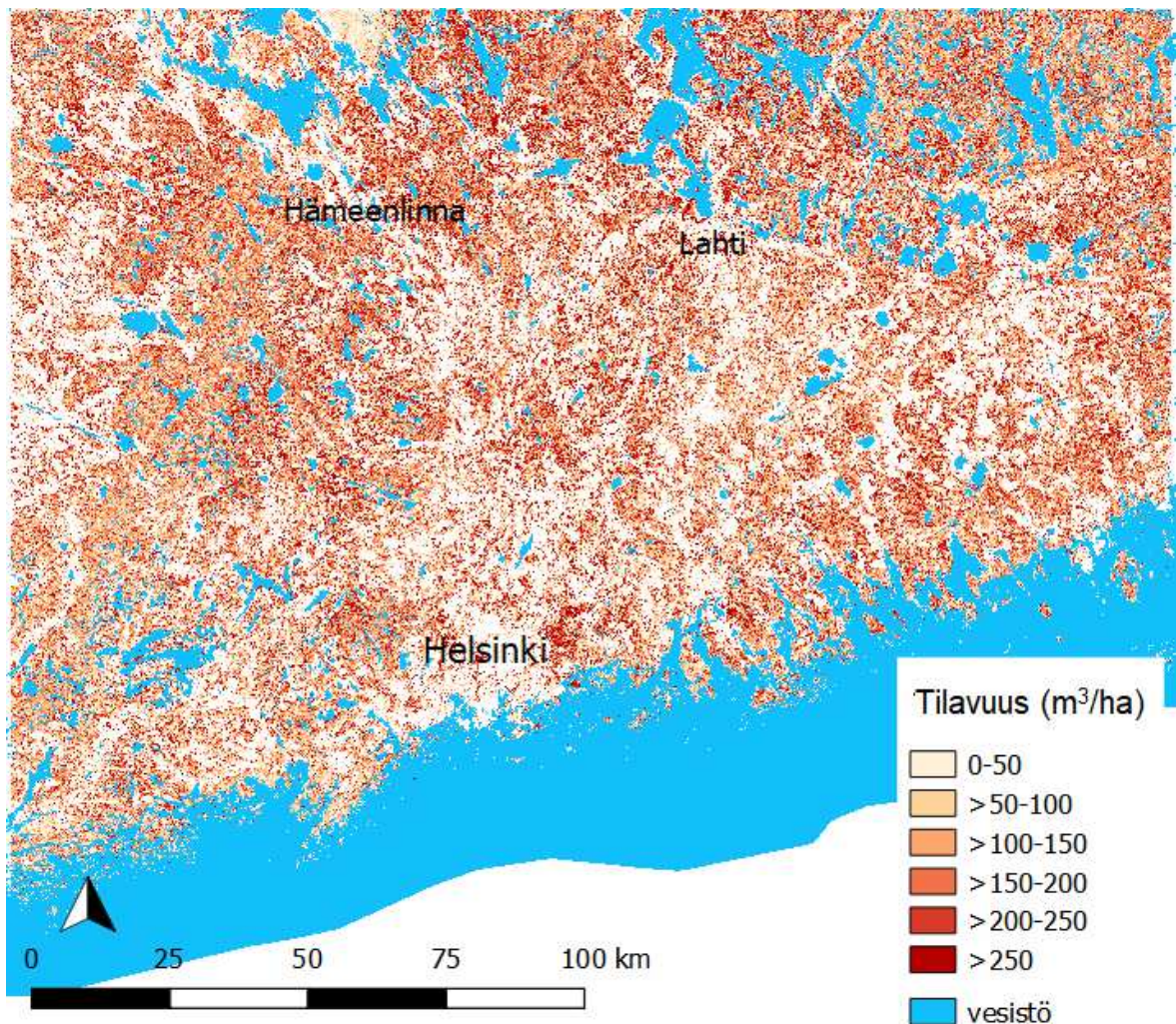
Tässä tutkimuksessa mallin validoinnin haasteena oli osassa puulajiluokissa havaintojen vähyys, joka voi selittää saatujen tunnuslukujen suurta vaihtelua. Vähemmän havaintoja sisältävillä puulajiluokilla validoinnissa saatujen tunnuslukujen suuruus vaihteli enemmän. Tupla-ristiinvalidoinnissa havainnot jaettiin kahteen ryhmään, jolloin lehtipuun, havusekametsän sekä sekametsän opetus- ja testiryhmiin jäi kumpaankin alle 80 havaintoa. Lehtipuulla nRMSE:n arvot vaihtelivat 39,9 prosenttiyksikköä, havusekametsällä 41,0 prosenttiyksikköä ja sekametsällä 25,6 prosenttiyksikköä. Vertailukohtana puulajit mänty, jolla nRMSE:n vaihtelu oli 9,9 prosenttiyksikköä ja kuusi, jolla vaihtelu oli 14,4 prosenttiyksikköä. Koko aineistolle lasketulla mallilla validoinnin tulokset olivat lähellä toisiaan ja

nRMSE vaihtelu oli 1,7 prosenttiyksikköä. Tämän perusteella voidaan todeta, että havaintojen lukumäärää kasvattamalla tilavuusarvioinnin tarkkuutta voitaisiin mahdollisesti parantaa myös puulajeilla, joissa havaintoja oli vähän.

Havaintojen lukumäärän kasvattamisen lisäksi tämän tutkimuksen tilavuusmallia voitaisiin parantaa tarkemmalla koeala-aineiston läpi käymisellä. Osa koealoista sijaitsivat lähellä tienreunaa tai muuta aluetta, joka poikkesi selvästi mitatusta koealasta. Yhteen pikseliin vaikuttaa myös naapuripikselien heijastus (Korhonen et al. 2017). Tämän seurauksena koealaa vastaava pikseli ei täysin edusta kohteen heijastusta vaan siihen on myös vaikuttanut viereisen kohteen heijastus. Suurempi vaikutus on selvästi koealasta poikkeavilla kohteilla. Mallin virheen vähentämiseksi koeala-aineistosta voitaisiin poistaa kaikki koealat, jotka sijaitsevat liian lähellä reuna-alueita.

#### 4.5 Tilavuuskartan vertailu MVMI-tuotteeseen

Puulajikohtaisia tilavuusmalleja sovellettiin koealoja vastaavilla Sentinel-2-tiilillä, jotka olivat luokiteltu puulajeittain Random Forest -luokittelun avulla. Tilavuusarvion laskemisen jälkeen tulos luokiteltiin kuuteen (6) luokkaan, joiden rajat olivat 0-50 m<sup>3</sup>/ha, >50-100 m<sup>3</sup>/ha, >100-150 m<sup>3</sup>/ha, >150-200 m<sup>3</sup>/ha, >200-250 m<sup>3</sup>/ha ja >250 m<sup>3</sup>/ha (Kuva 13).



Kuva 13. Tämän tutkimuksen puulajiluokakohtaisilla tilavuusmalleilla laskettu puuston tilavuus Etelä-Suomen alueella (61°17'N, 23°15'E ja 59°50'N, 27°10'E).



Kuvassa 13 on havaittavissa pilvisyyden aiheuttamat virheet tilavuusarviossa. Hämeenlinnan yläpuolella kuvan reunalla esiintyi pilvi (Kuva 14). Pilven alueella tilavuusarviot olivat pieniä (Kuva 14a). Nämä arviot eivät kuvaa puuston tilavuutta ja näin ollen ovat käyttökelvottomia.

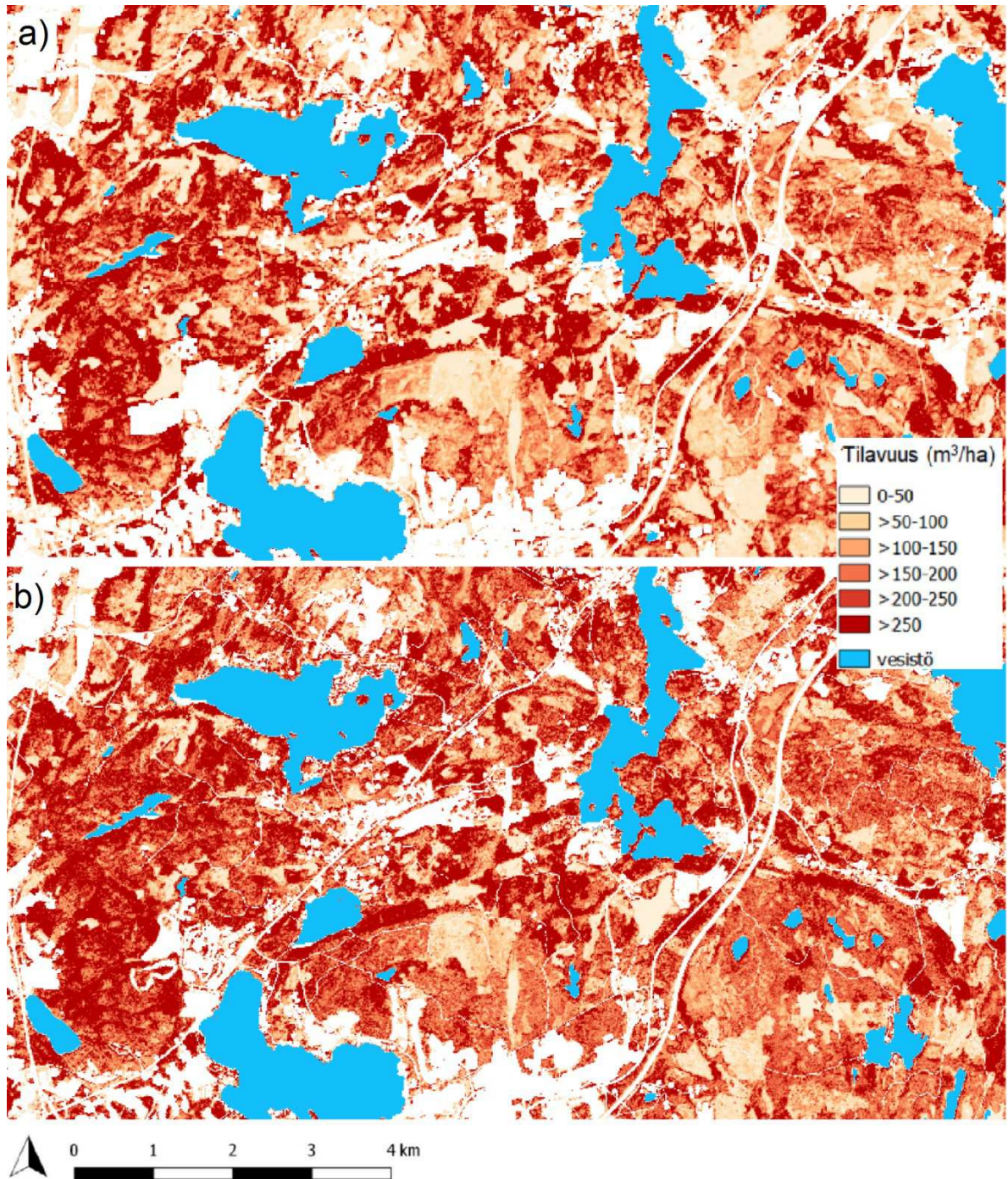


Kuva 14. a) Pilvisyyden aiheuttama virhe tilavuusarviossa. b) Alkuperäinen Sentinel-2-kuva (61°18' N, 24°13'E ja 61°10' N, 24°42'E).

Kuvasta 13 valittiin mahdollisimman metsäpeitteinen alue lähempään tarkasteluun (Kuva 15a). Alueella esiintyi sekä suuren että pienemmän tilavuuden metsiköitä, joten se antoi hyvän vertailukohdan MVMI-tuotteen tilavuusarvion kanssa. MVMI-tuote puuston tilavuudesta oli vuodelta 2015. Kuitenkaan MVMI-tuotteen vuoden 2015 tulosraporttia ei ollut tätä tutkimusta tehdessä vielä julkaistu aineistonlatauspalvelussa, joten varmaa aineiston keruujankohdasta ei ollut tiedossa. Ladattu MVMI-tuotteen tilavuusarvio luokiteltiin tätä tutkimustyötä varten kuuteen (6) edellä mainittuun tilavuusluokkaan (Kuva 15b), jotta tämän tutkimuksen sekä MVMI-tuotteen tilavuusarviot olivat vertailukelpoiset.

Visuaalisesti tarkasteltuna tämän tutkimustyön luokiteltu tilavuusarvio oli saman suuntainen MVMI-tuotteen tulosten kanssa (Kuva 15a ja b). Vertailemalla kuvia 15a ja b voitiin havaita, että tämän tutkimuksen tilavuuskartta näytti ”pehmeämmältä” kuin MVMI-tuotteen tulos. Karttakuvien väliset erot voivat osittain selittyä toisistaan eroavalla pikselikoolla. Tässä tutkimuksessa käytettyjen Sentinel-2-kanavien vuoksi pikselikoko oli 20 m, kun taas MVMI-tuotteen tulokset esitetään 16 m kokoisilla pikseleillä. Eniten eroavaisuuksia vaikutti olevan alueilla, joiden tilavuus oli 100-200 m<sup>3</sup>/ha. Korkean ja pienen tilavuuden alueet vastasivat hyvin toisiaan. Mallilla tuotetut arviot antoivat hyvin suuntaa tilavuuden suuruusluokasta. Tilavuusarvio näytti soveltuvan pikselikohtaista tarkastelua paremmin laajempien alueiden tilavuuden arvioimiseen.



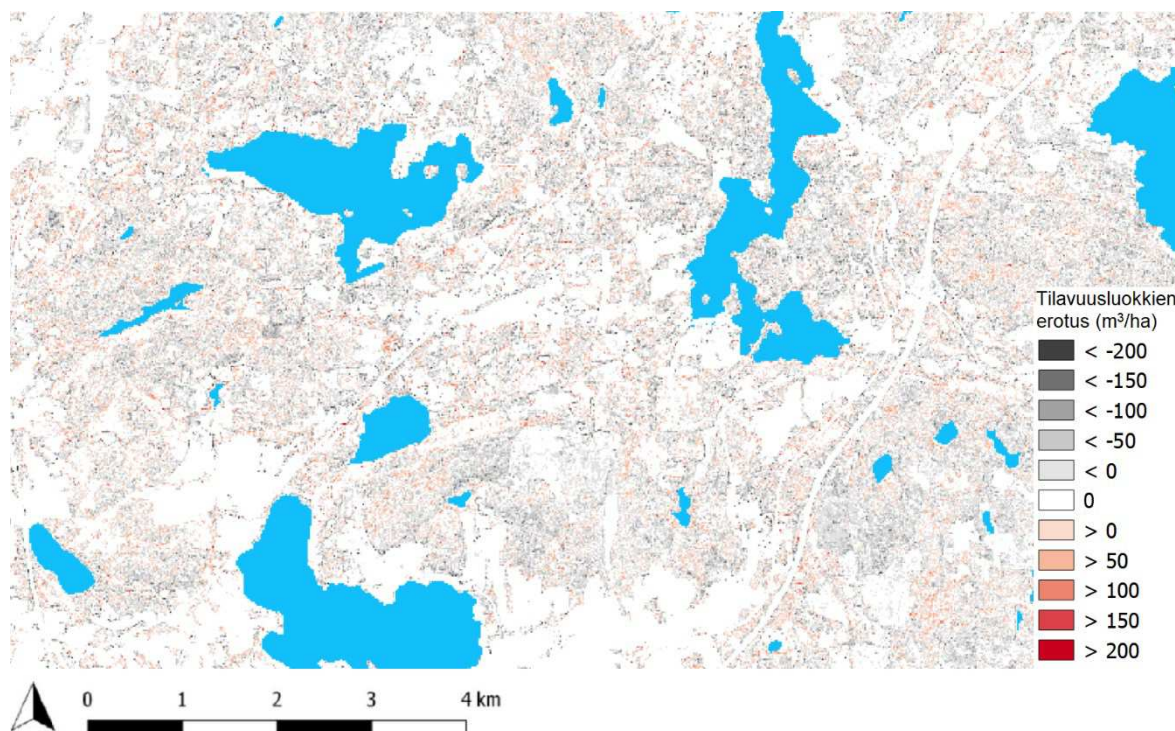


Kuva 15. a) Tämän tutkimuksen ja b) MVMI-tuotteen tilavuuskartta ( $61^{\circ}4'N$ ,  $25^{\circ}40'E$  ja  $61^{\circ}0'N$ ,  $25^{\circ}54'E$ ).

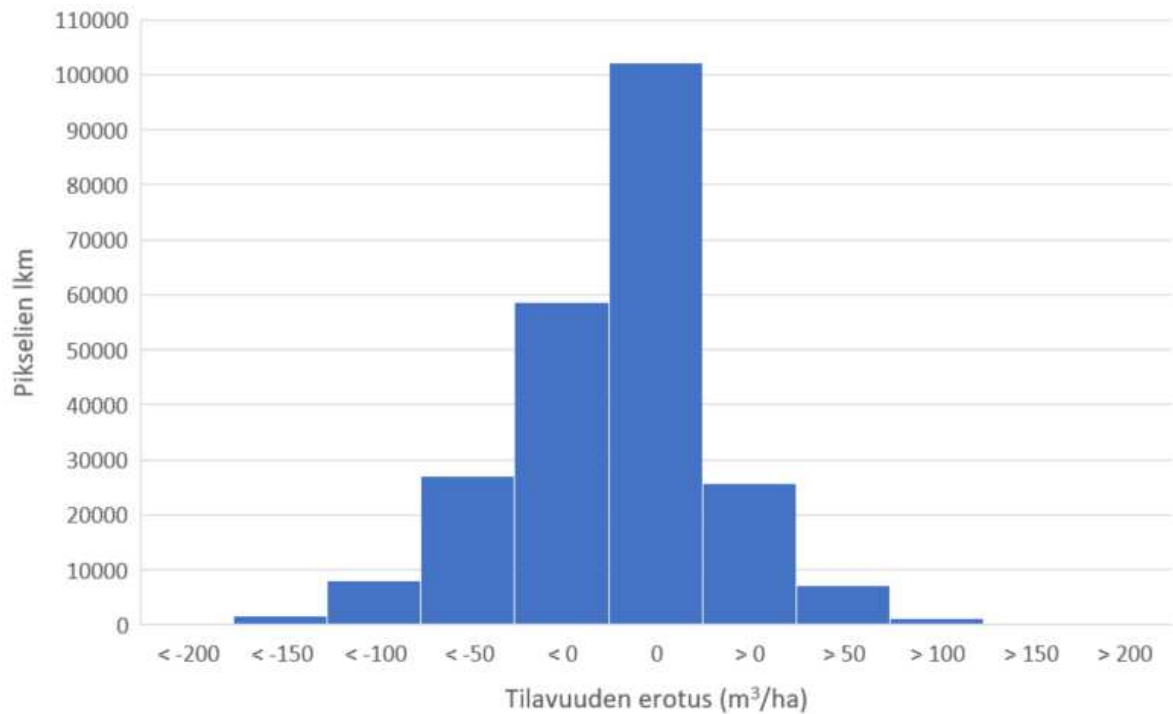
Visuaalisen tarkastelun lisäksi laskettiin erotuskartta (Kuva 16) ja sitä kuvaava histogrammi (Kuva 17) tämän tutkimuksen sekä MVMI-tuotteen luokitellun tilavuusarvion välillä. Erotuskartan negatiiviset arvot (harmaan sävyillä) tarkoittavat, että regressiomallin tilavuusarvio on pienempi kuin MVMI-tuotteen tilavuusarvio. Vastaavasti positiiviset arvot (punaisen sävyillä) tarkoittavat, että regressiomallin tilavuusarvio on MVMI-tuotteen tilavuusarviota suurempi.



Erotuskartalta voi havaita, että kuva oli pääasiassa harmaasävytteinen, joka näkyi myös histogrammista. Histogrammi oli vasemmalle vino, joka tarkoittaa, että tämän tutkimuksen tilavuusarviot olivat keskimäärin pienempiä kuin MVMI-tuotteen tilavuusarviot. Suurin osa tämän tutkimuksen tilavuusarviosta ja MVMI-tuotteen tilavuusarvioita oli luokiteltu samaan luokkaan, joka vastaa erotuskartalla valkoista väriä (0). Histogrammissa seuraavaksi eniten eroa kuvien välillä oli 0-(-50) m<sup>3</sup>/ha, joka tarkoittaa, että tämän tutkimuksen tilavuusarvio oli enintään 50 m<sup>3</sup>/ha pienempi kuin MVMI-tuotteen tilavuusarvio. Tilavuusarvioiden välillä ei esiintynyt suuria negatiivisia tai positiivisia eroja.

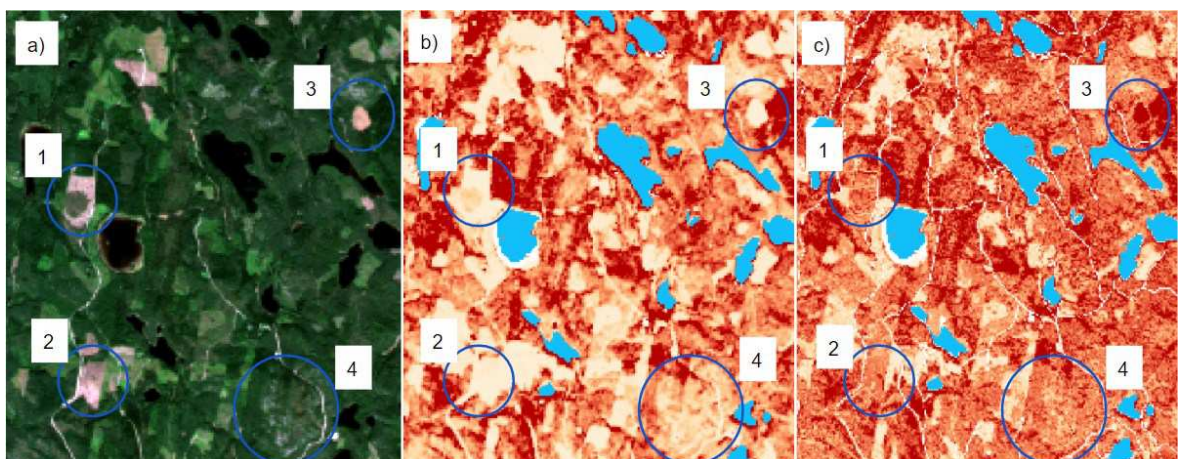


Kuva 16. Erotuskartta regressiomallilla lasketun luokitellun tilavuuskartan sekä MVMI-tuotteen luokitellun tilavuuskartan välillä. Negatiiviset arvot tarkoittavat, että regressiomallin tilavuuden tulokset olivat pienempiä kuin MVMI-tuotteen tilavuusarvio. Positiiviset arvot tarkoittavat, että regressiomallin tulokset olivat suurempia kuin MVMI-tuotteen tilavuusarvio.



Kuva 17. Erotuskartasta laskettu histogrammi, jossa esitetään tämän tutkimuksen regressiomallin ja MVMI-tuotteen tilavuusarvioiden erotuksien suuruus ja luokkaa vastaavien pikselien lukumäärä.

Eroavaisuuksia tilavuusarvioiden välillä voi myös aiheuttaa aineistojen keräämisen poikkeavat ajankohdat. Tämän tutkimuksen satelliittikuva oli elokuun vuoden 2015 kuva. MVMI-tuotteen tulos oli vuodelle 2015, mutta aineisto on oletettavasti kerätty aikaisemmin. Kuvasissa 18b ja c on havainnollistettu neljä poikkeavuutta MVMI-tuotteen ja tämän tutkimuksen tilavuusarvion välillä. Poimitut erot 1, 2, 3 ja 4 näkyivät satelliittikuvalla hakkuuaukoina tai harvapuustoisempina alueena, jolloin myös tämän tutkimuksen tilavuusarviot olivat näillä alueilla alhaisia (Kuva 18b). Vastaavasti MVMI-tuotteen tilavuusarvio oli näillä alueilla suurempi (Kuva 18c).



Kuva 18. a) Alkuperäinen Sentinel-2-kuva. b) Tämän tutkimuksen ja c) MVMI-tuotteen tilavuusarvio. Ympyröityjen alueiden (1-4) keskipisteiden koordinaatit 1: 61°11'N, 26°28'E, 2: 61°10'N, 26°29'E, 3: 61°11'N, 26°31'E, 4: 61°10'N, 26°31'E.



Eroavan ajankohdan lisäksi poikkeavuuksia karttojen välillä voi osittain selittää puulajiluokittelu. Tässä tutkimustyössä käytettyä puulajiluokittelua ei oltu optimoitu, joten parantamalla puulajiluokittelua myös tilavuusarvioinnin tarkkuus voisi parantua.

Pikselikohtaiset satunnaiset virheet voivat myös aiheutua dispersion eli hajaantumisen vaihtelusta ilmakehässä (Halme ja Tomppo 2001). Yhden pikselin heijastussuhdearvoon vaikuttaa puiden ja naapuripikselien lisäksi aluskasvillisuus, ojat sekä pienet pikselin sisälle sijoittuvat vesialueet. Esimerkiksi havupuiden latvuston heijastussuhde on matalampi kuin aluskasvillisuuden. (Korhonen et al. 2017.) Yksi mahdollisuus minimoida metsikön taustan heijastusta on käyttää kasvillisuusindeksejä (Rautiainen et al. 2006). Jatkotutkimuksessa voitaisiin selvittää parantaako Sentinel-2-kanavista lasketut kasvillisuusindeksit tilavuusarviota pohjoiseen havumetsävyöhykkeeseen kuuluvilla metsiköillä.

Muita jatkotutkimusaiheita voisi olla kuviokohtainen tilavuusarviointi Sentinel-2-kuvalla. Lopputuloksena syntyneen tilavuuskartan tilavuusarviot olivat kuvioiden reunoilla heikompia. Tämä aiheutuu reunapikseleistä, jotka eivät edusta selkeästi kumpaakaan kuviota, vaan heijastussuhdearvoon vaikuttaa molempien kuvioiden heijastus. Kuviokohtaisessa arvioinnissa suurempi painoarvo on kuvion keskellä sijaitsevilla pikseleillä (Mäkelä ja Pekkarinen 2004). Reunaa kohti mennessä pikselin painoarvo pienenee, jolloin reunapikseleiden vaikutus tilavuusarviossa vähenee.

## 5 Yhteenveto

Tässä tutkimuksessa selvitettiin avoimen satelliittiaineiston Sentinel-2-kuvan käyttöä puuston tilavuuden arvioinnissa pohjoisella havumetsävyöhykkeellä Etelä-Suomen alueella. Tilavuuden arviointiin soveltuva regressiomalli kehitettiin sekä koko koeala-aineistoon perustuen että jakamalla aineisto valtapuulajeittain seuraaviin luokkiin: mänty, kuusi, lehtipuu, havusekametsä ja sekametsä.

Maastomittauksissa metsäalueen tilavuuden arvioiminen perustuu koealamittauksiin, jossa metsikön puista mitataan tunnuksia. Mitattujen tunnuksien ja mallien avulla arvioidaan koko alueen puustoa. Satelliittikuvaan perustuvissa menetelmissä puuston arvioiminen pohjautuu sekä referenssiaineistoon että satelliittikuvaan. Referenssiaineiston tieto yhdistetään satelliittikuvan tietoon eri menetelmien kuten regressioanalyysin avulla. Lopputuloksena syntyneen mallin avulla voidaan arvioida puustoa koko kuvalla.

Tässä tutkimuksessa selvitettiin ensin tilavuusarviointiin soveltuvia kanavia Pearsonin korrelaatiokertoimien avulla. Sentinel-2-kanavat korreloivat jossain määrin keskenään ja useimmiten havaittavissa oli vähintään kaksi kanavaryhmää, joiden kanavat korreloivat keskenään voimakkaasti. Tilavuusarvioinnin menetelmänä oli lineaarinen regressioanalyysi, jossa parhaimmat tulokset saavutettiin, kun selitettävänä muuttujana käytettiin tilavuuden luonnollista logaritmia. Tilavuuden luonnollisen logaritmin kanssa parhaiten korreloi punaisen reunan ensimmäinen kanava (B5) kaikilla puulajiluokilla lukuun ottamatta havusekametsää, jolla parhaiten korreloiva kanava oli punaisen reunan toinen kanava (B6).

Tutkimuksen keskeisimpänä tuloksena havaittiin Sentinel-2 punaisen reunan kanavien soveltuvan hyvin tilavuusarviointiin. Multikollineaarisuuden aiheuttamien mallivirheiden välttämiseksi tilavuusarviointi onnistui luotettavasti enintään kahdella kanavalla. Männyn, kuusen, lehtipuun sekä sekametsän arvioinnissa paras tilavuusmalli syntyi punaisen reunan ensimmäisellä kanavalla (B5). Koko aineiston mallissa sekä puulajiluokassa havusekametsä paras malli muodostui käyttämällä kahta kanavaa. Koko aineiston mallissa kanavat olivat punaisen reunan ensimmäinen (B5) sekä toinen (B6) kanava ja havusekametsän mallissa vihreän valon aallonpituusalueen (B3) sekä punaisen reunan toinen (B6) kanava.

Havupuuta sisältävien puulajien tilavuusarviointi oli tarkempaa kuin lehtipuuta sisältävien. MVM-tuotteen tilavuusarvioinnin verrattuna tämän tutkimuksen arviot olivat jonkin verran pienempiä, mutta antoivat kuitenkin suuntaa tilavuuden suuruusluokasta. Kaiken kaikkiaan tuloksien tarkkuus ei ollut merkittävästi parempi tai huonompi verrattuna tässä tutkimuksessa tarkasteltuihin aikaisempiin tutkimustuloksiin.

## Lähdeluettelo

Asner, G. 1998. Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment*. Vol. 64: 3. S. 234-253. [Viitattu 19.2.2018]. DOI: 10.1016/S0034-4257(98)00014-5. ISSN 0034-4257.

Astola, H., Bounsaythip, C., Ahola, J., Häme, T., Parmes, E., Sirro, L. & Veikkanen, B. 2004. High-forest parameter estimation from high resolution remote sensing data. *International Society for Photogrammetry and Remote Sensing*. 20. kongressi. 12-23. heinäkuuta. Istanbul, Turkki. S. 355–340.

Atkinson, P. & Tatnall, A. 1997. Introduction Neural networks in remote sensing. *International Journal of Remote Sensing*. Vol 18:4. S. 699-709. [Viitattu 11.1.2018]. DOI: 10.1080/014311697218700. ISSN 0143-1161 (painettu) ISSN 1366-5901 (sähköinen).

Breiman, L. 2001. Random Forests. *Machine Learning*. Vol. 45:1. S. 5–32. [Viitattu 28.2.2018]. DOI: 10.1023/A:1010933404324. ISSN 1573-0565 (sähköinen).

Chrysafis, I., Mallinis, G., Siachalou, S. & Patias, P. 2017. Assessing the relationships between growing stock volume and Sentinel-2 imagery in a Mediterranean forest ecosystem. *Remote Sensing Letters*. Vol. 8:6. S. 508-517. [Viitattu 15.12.2017]. DOI: 10.1080/2150704X.2017.1295479. ISSN 2150-704X (painettu) ISSN 2150-7058 (sähköinen).

Daoud, J. 2017. Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*. Vol. 949, conference 1. 6 s. [Viitattu 27.2.2018]. DOI: 10.1088/1742-6596/949/1/012009. ISSN 1742-6588 (painettu) ISSN 1742-6596 (sähköinen).

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. & Bargellini, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*. Vol. 120. S. 25-36. [Viitattu 12.1.2018]. DOI: 10.1016/j.rse.2011.11.026. ISSN 0034-4257.

Eriksson, H., Eklundh, L., Kuusk., A. & Nilson, T. 2006. Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates. *Remote Sensing of Environment*. Vol. 103:4. S. 408-418. [Viitattu 19.2.2018]. DOI: 10.1016/j.rse.2006.04.005. ISSN 0034-4257.

ESA. 2017. SENTINEL-2 OPERATIONS. [Verkkosivu]. [Viitattu 23.10.2017]. Saatavissa: [http://www.esa.int/Our\\_Activities/Operations/Sentinel-2\\_operations](http://www.esa.int/Our_Activities/Operations/Sentinel-2_operations).

ESA Sentinel Online. 2017. SENTINEL-2 MSI Introduction. [Verkkosivu]. [Viitattu 23.10.2017]. Saatavissa: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi>.

Fraser, R. & Kaufman, Y. 1985. The Relative Importance of Aerosol Scattering and Absorption in Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. GE-23:5. S. 625-633. [Viitattu 3.3.2018]. DOI: 10.1109/TGRS.1985.289380. ISSN 0196-2892 (painettu) ISSN 1558-0644 (sähköinen).

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, B., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P. & Fernandez, V. 2017. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sensing*. Vol. 9:6, 584. 81 s. [Viitattu 21.2.2018]. DOI: 10.3390/rs9060584. ISSN 2072-4292.

Gascon, F., Cadau, E., Colin, O., Hoersch, B., Isola, C., López Fernández, B. & Martimort, P. 2014. Copernicus Sentinel-2 mission: products, algorithms and Cal/Val. *Proc. SPIE 9218, Earth Observing Systems XIX*. Vol. 9218:92181E. 10 s. [Viitattu 20.2.2018]. DOI: 10.1117/12.2062260.

Gatti, A. & Naud, C. 2017. Sentinel-2 Products Specification Document. ESA Sentinel 2 PDGS Core Producement. 27.9.2017. Issue: 14.3. 487 s. [Viitattu 22.2.2018].

Gschwantner, T., Schadauer, K., Vidal, C., Lanz, A., Tomppo, E., di Cosmo, L., Robert, N., Duursma, D. & Lawrence, M. 2009. Common tree definitions for national forest inventories in Europe. *Silva Fennica*. Vol. 43:2. S. 303–321. [Viitattu 23.2.2018]. Saatavissa: <http://www.metla.fi/silvafennica/full/sf43/sf432303.pdf>. ISSN 0037-5330 (painettu).

Grömping, U. 2012. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *The American Statistician*. Vol. 63:4. S. 308-319. [Viitattu 28.2.2018]. DOI: 10.1198/tast.2009.08199. ISSN 0003-1305 (painettu) 1537-2731 (sähköinen).

Günlü, A., Ercanlın, İ., Keleş, S. & Anlar, H. 2015. Modelling of stand volume and tree density using Spot-4 satellite image: a case study in Devrez planning unit. *International Journal of Global Warming*. Vol. 7:4. S. 454-465. [Viitattu 10.10.2017]. DOI: 10.1504/IJGW.2015.070047. ISSN 1758-2083 (painettu) ISSN 1758-2091 (sähköinen).

Hall, F., Huemmrich, K., Strebel, D., Goetz, S., Nickeson, J. & Woods, K. 1992. Biophysical, Morphological, Canopy Optical Property, and Productivity Data from the Superior National Forest. NASA Technical Memorandum 104568. NASA Goddard Space Flight Center: Greenbelt, MD, USA. 137 s. [Viitattu 19.2.2018]. Saatavissa: <https://ntrs.nasa.gov/search.jsp?R=19920020967>.

Hall, R., Skakun, R., Arsenault, B. & Case, B. 2006. Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. *Forest Ecology and Management*. Vol. 225:1-3. S. 378-390. [Viitattu 24.10.2017]. DOI: 10.1016/j.foreco.2006.01.014. ISSN 0378-1127.

Halme, M. & Tomppo, E. 2001. Improving the accuracy of multisource forest inventory estimates by reducing plot location error — a multicriteria approach. *Remote Sensing of Environment*. Vol. 78:3. S. 312-327. [Viitattu 15.11.2017]. DOI: 10.1016/S0034-4257(01)00227-9. ISSN 0034-4257.

Holmgren, J., Joyce, S., Nilsson, M. & Olsson, H. 2000. Estimating Stem Volume and Basal Area in Forest Compartments by Combining Satellite Image Data with Field Data. *Scandinavian Journal of Forest Research*. Vol. 15:1. S. 103-111. [Viitattu 25.10.2017]. DOI: 10.1080/02827580050160538. ISSN 0282-7581 (painettu) ISSN 1651-1891 (sähköinen).

Hovi, A., Raitio, P. & Rautiainen, M. 2017. A spectral analysis of 25 boreal tree species. *Silva Fennica*. Vol. 51:4. Article id 7753. [Viitattu 9.10.2017]. DOI: 10.14214/sf.7753. ISSN 2242-4075 (sähköinen).

Huiyan, G., Limin, D., Gang, W., Dong, X., Shunzhong, W. & Hui, W. 2006. Estimation of forest volumes by integrating Landsat TM imagery and forest inventory data. *Science in China Series E: Technological Sciences*. Vol. 49:(Suppl. 1). S. 54–62. [Viitattu 24.10.2017]. DOI: 10.1007/s11431-006-8107-z. ISSN 1862-281X (sähköinen).

Hyypä, J., Hyypä H., Inkinen, M., Engdahl, M., Linko, S., & Zhu, Y.-H. 2000. Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*. Vol. 128:1-2. S. 109-120. [Viitattu 24.10.2017]. DOI: 10.1016/S0378-1127(99)00278-9. ISSN 0378-1127.

Häme, T., Rauste, Y., Antropov, O., Ahola, H. & Kilpi, J. 2013. Improved Mapping of Tropical Forests With Optical and SAR Imagery, Part II: Above Ground Biomass Estimation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Vol. 6:1. S. 92-101. [Viitattu 21.2.2018]. DOI: 10.1109/JSTARS.2013.2241020. ISSN 1939-1404 (painettu) ISSN 2151-1535 (sähköinen).

Kershaw, J., Ducey, M., Beers, T. & Husch, B. 2017. *Forest mensuration*. 5. pain. Singapore. John Wiley & Sons. 613 s. ISBN 9781118902035.

Kilki, P. 1983. Sample trees in timber volume estimation. *Acta Forestalia Fennica* 182. Article id 7629. Suomen metsätieteellinen seura. S. 1-35. [Viitattu 3.3.2018]. Saatavissa: <https://doi.org/10.14214/aff.7629>. ISBN 951-651-057-4.

Kilki, P. & Päivinen, R. 1986. Weibull function in the estimation of the basal area dbh-distribution. *Silva Fennica*. Vol. 20:2. Article id 5270. S. 149-156. [Viitattu 12.3.2018] Saatavissa: <https://doi.org/10.14214/sf.a15449>. ISSN 2242-4075 (sähköinen).

Knipling, E. 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*. Vol. 1:3. S. 155-159. [Viitattu 3.3.2018]. DOI: 10.1016/S0034-4257(70)80021-9. ISSN 0034-4257.

Korhonen, L., Hadi, Packalen, P. & Rautiainen, M. 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. *Remote Sensing of Environment*. Vol. 195:15. S. 259-274. [Viitattu 21.12.2017]. DOI: 10.1016/j.rse.2017.03.021. ISSN 0034-4257.

Laasasenaho, J. 1982. Taper curve and volume functions for pine, spruce and birch. *Communicationes Instituti Forestalis Fenniae* 108. Helsinki, Suomi. Metsäntutkimuslaitos. 74 s. [Viitattu 10.3.2018]. Saatavissa: <http://urn.fi/URN:ISBN:951-40-0589-9>. ISBN 951-40-0589-9.

Laasasenaho, J. & Snellman, C.-G. 1983. Männyn, kuusen ja koivun tilavuustaulukot. Metsäntutkimuslaitoksen tiedonantoja 113. Helsinki, Suomi. Metsäntutkimuslaitos. 91 s. [Viitattu 23.2.2018]. Saatavissa: <http://urn.fi/URN:ISBN:951-40-0999-1>. ISBN 951-40-0999-1.

- Li, X. & Strahler, A. 1992. Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: effect of crown shape and mutual shadowing. *IEEE Transactions on Geoscience and Remote Sensing*. Vol. 30:2. S. 276-292. [Viitattu 3.3.2018]. DOI: 10.1109/36.134078. ISSN 0196-2892 (painettu) ISSN 1558-0644 (sähköinen).
- Lillesand, T., Kiefer, R. & Chipman, J. 2008. Remote sensing and image interpretation. 6. pain. United States of America. John Wiley & Sons. 756 s. ISBN 978-0-470-05245-7.
- Lehmonen, H. 2014. Metsäsuunnitelma ja puuston arviointi. Teoksessa: Rantala, S. (toim.) Metsäkoulu. 8. pain. Porvoo: Suomi. Metsäkustannus Oy. S. 229-244. ISBN 978-952-6612-27-0.
- Landsberg, J. 2003. Physiology in forest models: History and the future. *FBMIS*. Vol. 1. S. 49-63. [Viitattu 23.2.2018]. Saatavissa: [https://www.researchgate.net/publication/228906709\\_Physiology\\_in\\_forest\\_models\\_History\\_and\\_the\\_future](https://www.researchgate.net/publication/228906709_Physiology_in_forest_models_History_and_the_future). ISSN 1740-5955 (sähköinen).
- Lindberg, H. & Nygren, M. 2014. Metsänkasvatuksen lähtökohdat. Teoksessa: Rantala, S. (toim.) Metsäkoulu. 8. pain. Porvoo: Suomi. Metsäkustannus Oy. S. 21-56. ISBN 978-952-6612-27-0.
- Lindblad, J. 2014. Puutavaran mittaust. Teoksessa: Rantala, S. (toim.) Metsäkoulu. 8. pain. Porvoo: Suomi. Metsäkustannus Oy. S. 217-228. ISBN 978-952-6612-27-0.
- Lindsey, A., Barton, J. & Miles S. 1958. Field efficiencies of forest sampling methods. *Ecology*. Vol. 39:3. S. 428-444. [Viitattu 11.3.2018]. DOI: 10.2307/1931752. ISSN 1939-9170 (sähköinen).
- Lukeš, P., Stenberg, P., Rautiainen, M., Möttus, M. & Vanhatalo, K. 2013. Optical properties of leaves and needles for boreal tree species in Europe. *Remote Sensing Letters*. Vol. 4:7. S. 667-676. [Viitattu 19.2.2018]. DOI: 10.1080/2150704X.2013.782112. ISSN 2150-704X (painettu) ISSN 2150-7058 (sähköinen).
- Luonnonvarakeskus. 2017. Valtakunnan metsien inventointi: VMI1:n menetelmät ja tulostaulukot julkaistu. [Verkkouutinen]. 2.10.2017. [Viitattu 10.10.2017]. Saatavissa: <https://www.luke.fi/tietoa-luonnonvaroista/metsa/metsavarat-ja-metsasuunnittelu/metsavarat/>.
- Luonnonvarakeskus. 2012. Suomen metsät 2012: Kriteeri 4 Monimuotoisuus. [Verkkouutinen]. 21.3.2012. [Viitattu 8.11.2017]. Saatavissa: <http://www.metla.fi/metinfo/kestavyys/c4-tree-species.htm>.
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U. & Gascon, F. 2017. Sen2Cor for Sentinel-2. *Proc. SPIE 10427, Image and Signal Processing for Remote Sensing XXIII*. Vol. 10427:1042704. 13 s. [Viitattu 21.2.2018]. DOI: 10.1117/12.2278218.
- Mayer, B. & Kylling, A. 2005. Technical note: The libRadtran software package for radiative transfer calculations – description and examples of use. *Atmospheric Chemistry and*

Physics. Vol. 5:7. S. 1855–1877. [Viitattu 21.2.2018]. Saatavissa: <https://doi.org/10.5194/acp-5-1855-2005>. ISSN 1680-7324 (sähköinen).

Mohammadi, J., Joibary, S., Yaghmaee, F. & Mahiny, A. 2010. Modelling forest stand volume and tree density using Landsat ETM+ data. *International Journal of Remote Sensing*. Vol. 31:11. S. 2959-2975. [Viitattu 25.10.2017]. DOI: 10.1080/01431160903140811. ISSN 0143-1161 (painettu) ISSN 1366-5901 (sähköinen).

Metsäkeskus. 2014. Kaukokartoitusperusteisen metsien inventoinnin koealojen maastotyöohje. Maastotyöohje versio 1.4. 21.5.2014. 32 s.

Metsäkeskus. 2016. Metsän mitat ja yksiköt. [Verkkosivu]. [Viitattu 25.10.2017]. Saatavissa: <https://www.metsakeskus.fi/metsan-mitat-ja-yksikot>.

Moeur, M. 1988. Nearest neighbour inference for correlated multivariate attributes. USDA Forest Service General Technical Report. Issue 120. S. 716–723. ISSN 0363-616X.

Montgomery, D., Peck, E. & Vining, G. 2012. Introduction to linear regression analysis. 5. pain. New Jersey, USA. John Wiley & Sons Ltd. 679 s. (Wiley series in probability and statistics). ISBN 9781119180173 (sähköinen).

Möttö, M., Sulev, M. & Hallik, L. 2014. Seasonal course of the spectral properties of alder and birch leaves. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Vol. 7:6. S. 2496–2505. [Viitattu 3.3.2018]. DOI: 10.1109/JSTARS.2013.2294242. ISSN 1939-1404 (painettu) ISSN 2151-1535 (sähköinen).

Mäkelä, H. & Pekkarinen, A. 2004. Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*. Vol. 196:2-3. S. 245-255. [Viitattu: 24.10.2017]. DOI: 10.1016/j.foreco.2004.02.049. ISSN 0378-1127.

Nyyssönen, A. & Mielikäinen, K. 1978. Metsikön kasvun arviointi. *Acta Forestalia Fennica* 163. Article id 7597. Suomen metsätieteellinen seura. 40 s. [Viitattu 10.3.2018]. Saatavissa: <https://doi.org/10.14214/aff.7597>. ISBN 951-651-038-8.

Palosuo, V. 1979. MERA-ohjelmat Suomen metsätaloudessa. *Acta Forestalia Fennica* 165. Article id 7599. 55 s. Saatavissa: <https://doi.org/10.14214/aff.7599>. ISBN 951-651-040-X.

Peuhkurinen, J., Maltamo, M., Vesa, L. & Packalén, P. 2008. Estimation of Forest Stand Characteristics Using Spectral Histograms Derived from an Ikonos Satellite Image. *Photogrammetric Engineering & Remote Sensing*. Vol. 74:11, S. 1335-1341. [Viitattu 24.10.2017]. DOI: 10.14358/PERS.74.11.1335. ISSN 0099-1112 (painettu).

Poso, S. 1983. Kuvioittaisen arvioimismenetelmän perusteita. *Silva Fennica*. Vol. 17:4. Article id 5196. S. 313-349. [Viitattu 23.2.2018]. Saatavissa: <https://doi.org/10.14214/sf.a15179>. ISSN 2242-4075 (sähköinen).

Päivinen, R. 1980. Puiden läpimittajakauman estimointi ja siihen perustuva puustotunnusten laskenta. *Folia Forestalia* 442. 28 s. [Viitattu 12.3.2018]. Saatavissa: <http://urn.fi/URN:ISBN:951-40-0461-2>. ISBN 951-40-0461-2.

Päivinen, R., Nousiainen, M. & Korhonen, K. 1992. Puutunnusten mittaamisen luotettavuus. *Folia Forestalia* 787. 18 s. [Viitattu 3.3.2018]. Saatavissa: <http://urn.fi/URN:ISBN:951-40-1197-X>. ISBN 951-40-1197-X.

Rautiainen, M. & Lukeš, P. 2015. Spectral contribution of understory to forest reflectance in a boreal site: an analysis of EO-1 Hyperion data. *Remote Sensing of Environment*. Vol. 171. S. 98–104. [Viitattu 19.2.2018]. DOI: 10.1016/j.rse.2015.10.009. ISSN 0034-4257.

Rautiainen, M., Lukeš, P., Homolová, L., Hovi, A., Pisek, J. & Möttus, M. 2018. Spectral Properties of Coniferous Forests: A Review of In Situ and Laboratory Measurements. *Remote Sensing*. Vol. 10:2. 28 s. [Viitattu 12.2.2018]. DOI: 10.3390/rs10020207. ISSN 2072-4292.

Rautiainen, M., Stenberg, P., Heiskanen, J., Möttus, M., Korhonen, L., Peltoniemi, J., Suomalainen, J., Kaasalainen, S. & Manninen T. 2008. Metsän kaukokartoituksen perustutkimus? *Metsätieteen aikakauskirja vuosikerta 2008*. Nro 2 artikkeli id 6617. [Viitattu 15.1.2018]. DOI: 10.14214/ma.6617. ISSN 2489-3188.

Schaepman-Strub, G., Schaepman, M., Painter, T., Dangel, S. & Martonchik, J. 2006. *Remote Sensing of Environment*. Vol. 103:1. S. 27-42. [Viitattu 12.2.2018]. DOI: 10.1016/j.rse.2006.03.002. ISSN 0034-4257.

Tomppo, E. & Katila, M. 1991. Satellite image-based national forest inventory of Finland for publication in the igrss' 91 digest. *Geoscience and Remote Sensing Symposium, 1991. IGARSS '91. Remote Sensing: Global Monitoring for Earth Management., International*. S. 1141-1144. [Viitattu 10.3.2018]. DOI: 10.1109/IGARSS.1991.579272. ISBN: 0-87942-675-6.

Tomppo, E., Olsson H., Ståhl, G., Nilsson M., Hagner, O. & Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sensing of Environment*. Vol. 112:5. S. 1982-1999. [Viitattu 23.2.2018]. DOI: 10.1016/j.rse.2007.03.032. ISSN 0034-4257.

Trotter, C., Dymond, J. & Goulding, C. 1997. Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal of Remote Sensing*. Vol. 18:10. S. 2209-2223. [Viitattu 24.10.2017]. DOI: 10.1080/014311697217846. ISSN 0143-1161 (painettu) ISSN 1366-5901 (sähköinen).

Wulder, M. 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*. Vol. 22:4. S. 449-476. [Viitattu 16.2.2018]. DOI: 10.1177/030913339802200402. ISSN 03091333.

Yan, X. & Su, X. 2009. *Linear regression analysis: theory and computing*. Singapore: World Scientific. 328 s. ISBN 978-981-283-410-2.



## Liiteluettelo

Liite 1. Pearsonin korrelaatiokertoimen tulokset puulajiluokittain Sentinel-2-kanavien ja puuston tilavuuden sekä tilavuuden luonnollisen logaritmin välillä. 3 sivua.

Liite 2. Regressioanalyysin tulokset puulajiluokittain. 2 sivua.

Liite 3. Kuvaajat arvioidun ja mitatun tilavuuden välillä puulajiluokittain. 6 sivua.

Liite 4. Regressioanalyysin validoinnin tulokset puulajiluokittain. 2 sivu.

## Liite 1. Pearsonin korrelaatiokertoimen tulokset puulaji-luokittain Sentinel-2-kanavien ja puuston tilavuuden sekä tilavuuden luonnollisen logaritmin välillä

*Taulukko 1. Koko aineiston korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Koko ai-neisto	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,88	0,89	0,80	0,52	0,45	0,46	0,45	0,74	0,81
B3	0,88	1,00	0,90	0,96	0,74	0,66	0,67	0,67	0,86	0,87
B4	0,89	0,90	1,00	0,87	0,50	0,42	0,42	0,43	0,81	0,91
B5	0,80	0,96	0,87	1,00	0,79	0,72	0,71	0,73	0,91	0,89
B6	0,52	0,74	0,50	0,79	1,00	0,99	0,96	0,99	0,81	0,65
B7	0,45	0,66	0,42	0,72	0,99	1,00	0,96	1,00	0,77	0,59
B8	0,46	0,67	0,42	0,71	0,96	0,96	1,00	0,97	0,75	0,58
B8a	0,45	0,67	0,43	0,73	0,99	1,00	0,97	1,00	0,78	0,60
B11	0,74	0,86	0,81	0,91	0,81	0,77	0,75	0,78	1,00	0,95
B12	0,81	0,87	0,91	0,89	0,65	0,59	0,58	0,60	0,95	1,00
V	-0,52	-0,67	-0,54	-0,70	-0,65	-0,61	-0,63	-0,62	-0,64	-0,59
lnV	-0,61	-0,81	-0,65	-0,84	-0,71	-0,65	-0,68	-0,67	-0,74	-0,70

*Taulukko 2. Puulajiluokan männyn korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Mänty	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,87	0,89	0,77	0,26	0,18	0,23	0,20	0,76	0,81
B3	0,87	1,00	0,89	0,95	0,54	0,44	0,46	0,46	0,87	0,85
B4	0,89	0,89	1,00	0,86	0,23	0,13	0,14	0,16	0,89	0,95
B5	0,77	0,95	0,86	1,00	0,59	0,48	0,49	0,52	0,92	0,87
B6	0,26	0,54	0,23	0,59	1,00	0,98	0,94	0,99	0,47	0,27
B7	0,18	0,44	0,13	0,48	0,98	1,00	0,94	0,99	0,37	0,17
B8	0,23	0,46	0,14	0,49	0,94	0,94	1,00	0,94	0,38	0,19
B8a	0,20	0,46	0,16	0,52	0,99	0,99	0,94	1,00	0,42	0,21
B11	0,76	0,87	0,89	0,92	0,47	0,37	0,38	0,42	1,00	0,96
B12	0,81	0,85	0,95	0,87	0,27	0,17	0,19	0,21	0,96	1,00
V	-0,42	-0,59	-0,47	-0,65	-0,53	-0,47	-0,52	-0,49	-0,56	-0,48
lnV	-0,48	-0,69	-0,49	-0,75	-0,68	-0,60	-0,66	-0,62	-0,63	-0,52

*Taulukko 3. Puulajiluokan kuusen korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Kuusi	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,95	0,94	0,90	0,82	0,79	0,77	0,79	0,86	0,88
B3	0,95	1,00	0,94	0,98	0,91	0,88	0,87	0,88	0,92	0,91
B4	0,94	0,94	1,00	0,92	0,78	0,75	0,73	0,75	0,89	0,93
B5	0,90	0,98	0,92	1,00	0,93	0,90	0,88	0,91	0,96	0,93
B6	0,82	0,91	0,78	0,93	1,00	0,99	0,98	0,99	0,90	0,82
B7	0,79	0,88	0,75	0,90	0,99	1,00	0,97	1,00	0,87	0,79
B8	0,77	0,87	0,73	0,88	0,98	0,97	1,00	0,98	0,85	0,76
B8a	0,79	0,88	0,75	0,91	0,99	1,00	0,98	1,00	0,88	0,79
B11	0,86	0,92	0,89	0,96	0,90	0,87	0,85	0,88	1,00	0,97
B12	0,88	0,91	0,93	0,93	0,82	0,79	0,76	0,79	0,97	1,00
V	-0,64	-0,73	-0,60	-0,74	-0,76	-0,76	-0,75	-0,76	-0,68	-0,63
lnV	-0,79	-0,89	-0,79	-0,92	-0,89	-0,86	-0,86	-0,87	-0,89	-0,83

*Taulukko 4. Puulajiluokan lehtipuun korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Lehtipuu	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,86	0,81	0,78	0,43	0,29	0,34	0,31	0,69	0,74
B3	0,86	1,00	0,89	0,96	0,55	0,36	0,41	0,39	0,79	0,84
B4	0,81	0,89	1,00	0,91	0,34	0,17	0,19	0,20	0,76	0,87
B5	0,78	0,96	0,91	1,00	0,58	0,40	0,41	0,43	0,85	0,89
B6	0,43	0,55	0,34	0,58	1,00	0,96	0,87	0,97	0,74	0,55
B7	0,29	0,36	0,17	0,40	0,96	1,00	0,88	0,99	0,63	0,41
B8	0,34	0,41	0,19	0,41	0,87	0,88	1,00	0,89	0,61	0,43
B8a	0,31	0,39	0,20	0,43	0,97	0,99	0,89	1,00	0,66	0,45
B11	0,69	0,79	0,76	0,85	0,74	0,63	0,61	0,66	1,00	0,94
B12	0,74	0,84	0,87	0,89	0,55	0,41	0,43	0,45	0,94	1,00
V	-0,35	-0,47	-0,41	-0,48	-0,19	-0,10	-0,26	-0,14	-0,41	-0,46
lnV	-0,61	-0,77	-0,69	-0,77	-0,32	-0,16	-0,29	-0,20	-0,60	-0,67

*Taulukko 5. Puulajiluokan havusekametsän korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Havu-seka-metsä	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,84	0,78	0,74	0,65	0,56	0,61	0,57	0,71	0,71
B3	0,84	1,00	0,93	0,96	0,87	0,79	0,82	0,80	0,88	0,87
B4	0,78	0,93	1,00	0,91	0,76	0,66	0,69	0,68	0,89	0,92
B5	0,74	0,96	0,91	1,00	0,91	0,85	0,85	0,87	0,91	0,89
B6	0,65	0,87	0,76	0,91	1,00	0,98	0,97	0,98	0,82	0,74
B7	0,56	0,79	0,66	0,85	0,98	1,00	0,96	0,99	0,75	0,65
B8	0,61	0,82	0,69	0,85	0,97	0,96	1,00	0,96	0,74	0,66
B8a	0,57	0,80	0,68	0,87	0,98	0,99	0,96	1,00	0,78	0,69
B11	0,71	0,88	0,89	0,91	0,82	0,75	0,74	0,78	1,00	0,98
B12	0,71	0,87	0,92	0,89	0,74	0,65	0,66	0,69	0,98	1,00
V	-0,45	-0,69	-0,56	-0,72	-0,80	-0,81	-0,79	-0,80	-0,58	-0,51
lnV	-0,59	-0,81	-0,67	-0,82	-0,85	-0,83	-0,83	-0,83	-0,67	-0,61

*Taulukko 6. Puulajiluokan sekametsän korrelaatiokertoimet Sentinel-2-kanavien ja tilavuuden sekä tilavuuden luonnollisen logaritmin välillä.*

Seka- metsä	B2	B3	B4	B5	B6	B7	B8	B8a	B11	B12
B2	1,00	0,88	0,90	0,79	0,51	0,44	0,42	0,44	0,79	0,84
B3	0,88	1,00	0,91	0,96	0,72	0,65	0,61	0,65	0,90	0,90
B4	0,90	0,91	1,00	0,88	0,49	0,42	0,40	0,42	0,88	0,95
B5	0,79	0,96	0,88	1,00	0,75	0,68	0,64	0,69	0,94	0,91
B6	0,51	0,72	0,49	0,75	1,00	0,98	0,94	0,99	0,74	0,60
B7	0,44	0,65	0,42	0,68	0,98	1,00	0,95	0,99	0,69	0,53
B8	0,42	0,61	0,40	0,64	0,94	0,95	1,00	0,95	0,66	0,51
B8a	0,44	0,65	0,42	0,69	0,99	0,99	0,95	1,00	0,70	0,54
B11	0,79	0,90	0,88	0,94	0,74	0,69	0,66	0,70	1,00	0,97
B12	0,84	0,90	0,95	0,91	0,60	0,53	0,51	0,54	0,97	1,00
V	-0,42	-0,61	-0,49	-0,66	-0,62	-0,59	-0,58	-0,60	-0,61	-0,55
lnV	-0,54	-0,80	-0,68	-0,85	-0,70	-0,65	-0,60	-0,66	-0,78	-0,76

## Liite 2. Regressioanalyysin tulokset puulajiluokittain

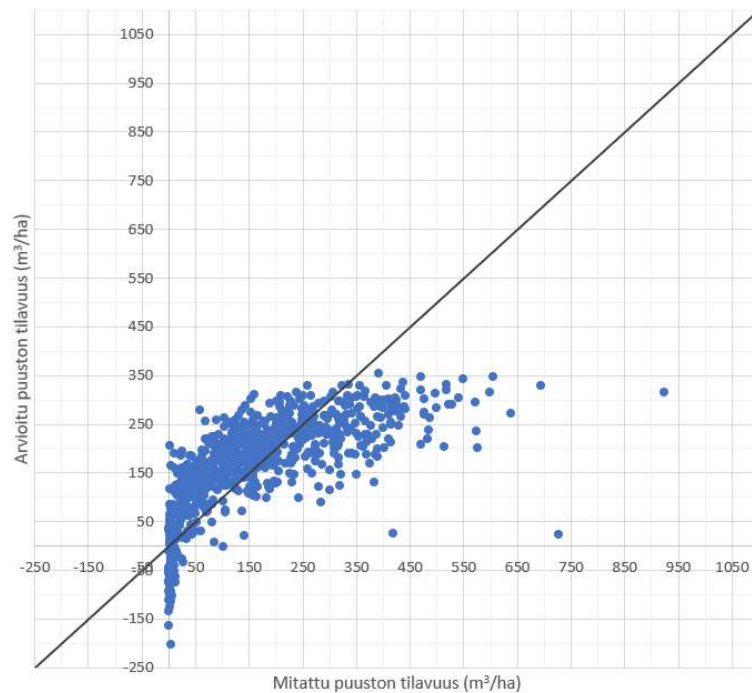
*Taulukko 1. Regressioanalyysin tulokset, kun selitettävänä muuttujana oli tilavuus.*

Puulaji	Kanavat	Korjattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MedAE (m <sup>3</sup> /ha)	nMedAE (%)	Mallin maksimi (m <sup>3</sup> /ha)	Mallin minimi (m <sup>3</sup> /ha)
Koko ai- neisto	B5	0,48	100,0	61,6	64,6	47,3	326,4	-220,1
	B3	0,44	103,7	63,9	67,0	49,1	357,2	-251,6
	B6	0,42	105,9	65,3	63,0	46,1	351,5	-99,4
	B3, B6	0,50	98,5	60,7	60,9	44,6	363,0	-204,3
	B5, B6	0,51	97,4	60,0	60,2	44,1	354,4	-202,9
	B3, B11	0,46	101,9	62,8	65,5	48,0	342,4	-246,4
Mänty	B6, B11	0,46	102,3	63,0	62,3	45,6	353,1	-130,2
	B5	0,42	70,6	46,6	44,7	30,5	259,3	-67,1
	B3	0,34	75,4	49,7	51,0	34,7	249,7	-73,3
	B11	0,31	76,9	50,7	51,4	35,1	241,5	-85,6
	B3, B11	0,35	74,7	49,3	47,4	32,3	251,0	-66,9
	B5, B6	0,45	68,5	45,1	44,4	30,3	278,5	-77,5
Kuusi	B3, B6	0,40	71,4	47,1	46,9	32,0	284,7	-84,6
	B6, B11	0,40	71,5	47,1	44,5	30,3	288,3	-51,5
	B6	0,58	109,7	49,9	71,1	35,5	417,0	-137,4
	B5	0,54	114,9	52,3	82,7	41,2	365,8	-160,8
	B7	0,57	110,6	50,3	72,4	36,1	432,3	-125,0
	B8a	0,57	110,6	50,3	73,8	36,8	415,2	-129,2
Lehtipuu	B8	0,56	112,2	51,1	72,5	36,2	435,0	-192,9
	B5, B8	0,59	108,7	49,5	73,7	36,7	404,3	-173,9
	B3, B7	0,59	108,6	49,4	70,5	35,2	413,0	-117,7
	B3, B8a	0,59	108,7	49,5	73,9	36,9	402,4	-135,2
	B3, B8	0,58	109,4	49,8	74,1	37,0	381,7	-134,7
	B5	0,23	93,0	98,3	44,6	65,6	174,6	-95,0
Havuse- kametsä	B3	0,21	94,1	99,4	44,4	65,4	178,6	-66,6
	B12	0,20	94,6	100,0	48,2	70,9	184,7	-91,4
	B3, B12	0,22	93,3	98,6	45,4	66,9	181,0	-62,9
	B7	0,64	78,8	41,3	54,9	31,4	362,0	-91,5
	B6	0,63	115,5	60,5	82,4	47,0	295,2	-203,2
	B8a	0,64	79,5	41,7	54,9	31,4	360,6	-130,5
Seka- metsä	B5	0,44	104,6	70,6	79,0	72,3	280,9	-164,9
	B6	0,38	109,8	74,1	68,5	62,6	337,5	-56,0
	B3	0,37	110,3	74,4	78,6	71,8	279,4	-215,7
	B5, B6	0,47	101,5	68,5	65,9	60,3	323,4	-157,7
	B3, B6	0,43	104,5	70,5	64,6	59,1	335,0	-179,7
	B6, B11	0,43	105,1	70,9	67,3	61,6	336,9	-143,8

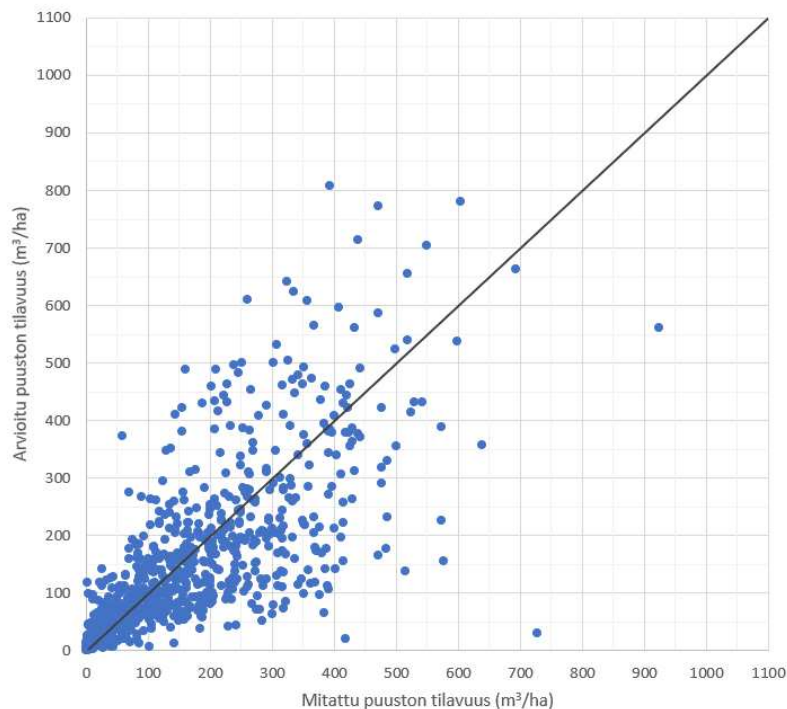
*Taulukko 2. Regressioanalyysin tulokset, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi.*

Puulaji	Kanavat	Korjattu $R^2$	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MedAE (m <sup>3</sup> /ha)	nMedAE (%)	Mallin maksimi (m <sup>3</sup> /ha)	Mallin minimi (m <sup>3</sup> /ha)
Koko ai- neisto	B5	0,71	104,0	64,0	39,3	28,8	715,9	0,6
	B3	0,65	105,8	65,2	44,2	32,4	1069,9	0,4
	B11	0,55	110,5	68,1	52,0	38,1	628,6	0,8
	B3, B11	0,67	105,8	65,2	44,5	32,6	955,5	0,4
	B5, B6	0,72	102,2	63,0	38,2	28,0	807,1	0,7
Mänty	B3, B6	0,68	110,7	68,2	39,6	29,0	1027,8	0,6
	B5	0,55	75,6	49,8	37,7	25,7	471,9	5,5
	B3	0,48	79,8	52,6	43,0	29,3	430,8	4,6
	B6	0,46	89,9	59,3	48,9	33,3	590,4	3,9
	B5, B8	0,67	103,2	68,0	40,9	27,9	967,9	4,2
	B5, B6	0,64	85,7	56,5	40,0	27,3	709,0	4,5
Kuusi	B3, B6	0,61	90,7	59,8	40,0	27,3	778,7	3,4
	B5	0,84	103,8	47,2	52,2	26,0	610,7	1,4
	B3	0,79	113,1	51,5	55,3	27,6	886,4	0,6
	B6	0,78	134,8	61,3	55,8	27,8	945,9	2,5
	B11	0,79	127,1	57,8	68,9	34,3	661,8	1,5
	B3, B8a	0,82	115,4	52,5	48,5	24,2	796,6	1,6
	B8a, B11	0,82	124,8	56,8	54,7	27,3	844,6	1,5
Lehtipuu	B7, B11	0,82	125,7	57,2	55,7	27,8	839,7	1,5
	B3, B7	0,81	115,1	52,4	45,2	22,6	808,8	1,6
	B5	0,59	97,7	103,2	20,3	29,8	260,5	0,7
	B3	0,59	99,7	105,4	25,4	37,4	308,4	1,1
	B4	0,47	103,2	109,1	29,2	43,0	167,7	1,1
Havuse- kametsä	B4, B12	0,49	100,8	106,5	28,8	42,5	197,2	1,0
	B6	0,72	97,3	51,0	42,4	24,2	707,9	3,8
	B7	0,68	93,6	49,1	41,4	23,6	628,9	7,5
	B8	0,69	109,6	57,5	44,5	25,4	704,4	6,3
	B8a	0,69	95,6	50,1	36,3	20,7	633,5	5,0
	B3, B6	0,73	93,5	49,0	30,2	17,2	710,0	3,0
	B3, B7	0,74	94,2	49,4	38,1	21,7	732,5	3,5
	B5, B7	0,72	95,6	50,1	31,9	18,2	734,4	4,2
Seka- metsä	B5, B8	0,73	98,2	51,5	36,1	20,6	786,6	3,9
	B5	0,72	115,0	77,6	38,1	34,8	552,9	0,4
	B3	0,64	114,0	76,9	42,4	38,7	560,8	0,4
	B11	0,60	137,0	92,4	40,6	37,1	781,1	0,1
	B3, B6	0,65	117,6	79,4	36,2	33,1	724,4	0,3

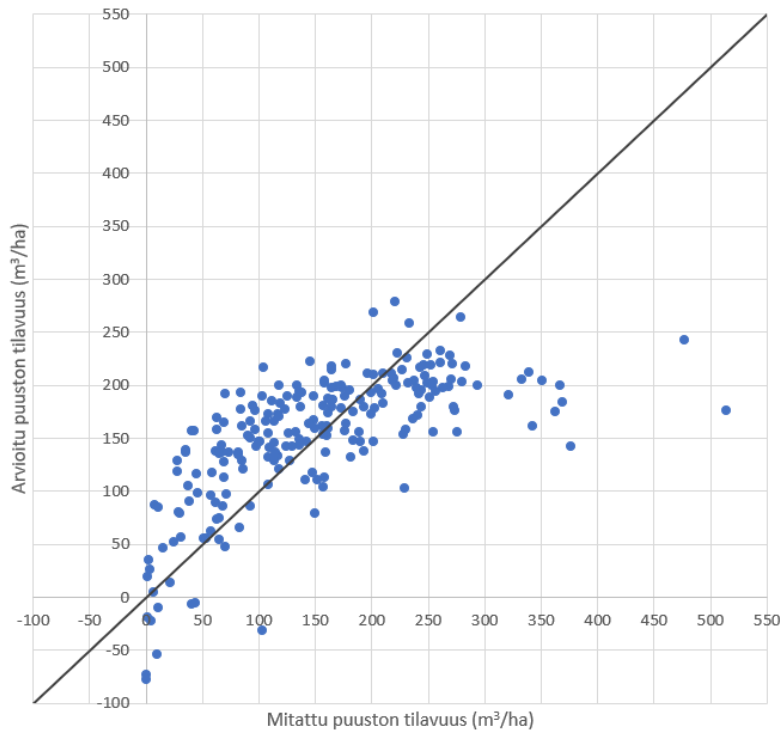
### Liite 3. Kuvaajat arvioidun ja mitatun tilavuuden välillä puulajiluokittain



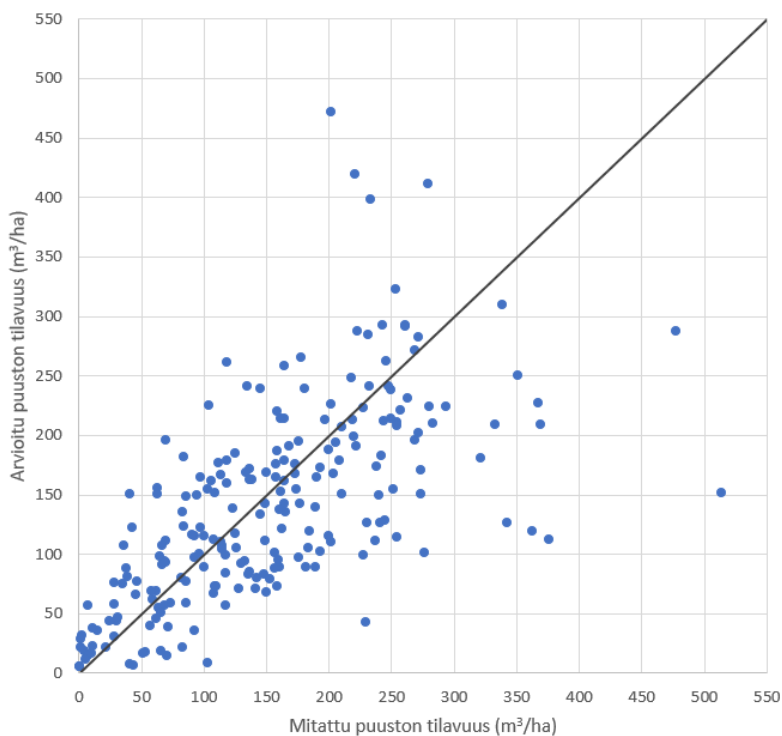
Kuva 1. Koko aineiston regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,51, RMSE 97,4  $m^3/ha$ , nRMSE 60,0 %, MdAE 60,2  $m^3/ha$ , nMdAE 44,1 %, maksimitilavuus 354,4  $m^3/ha$  sekä minimitilavuus -202,9  $m^3/ha$ .



Kuva 2. Koko aineiston regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,72, RMSE 102,2  $m^3/ha$ , nRMSE 63,0 %, MdAE 38,2  $m^3/ha$ , nMdAE 28,0 %, maksimitilavuus 807,1  $m^3/ha$  sekä minimitilavuus 0,7  $m^3/ha$ .

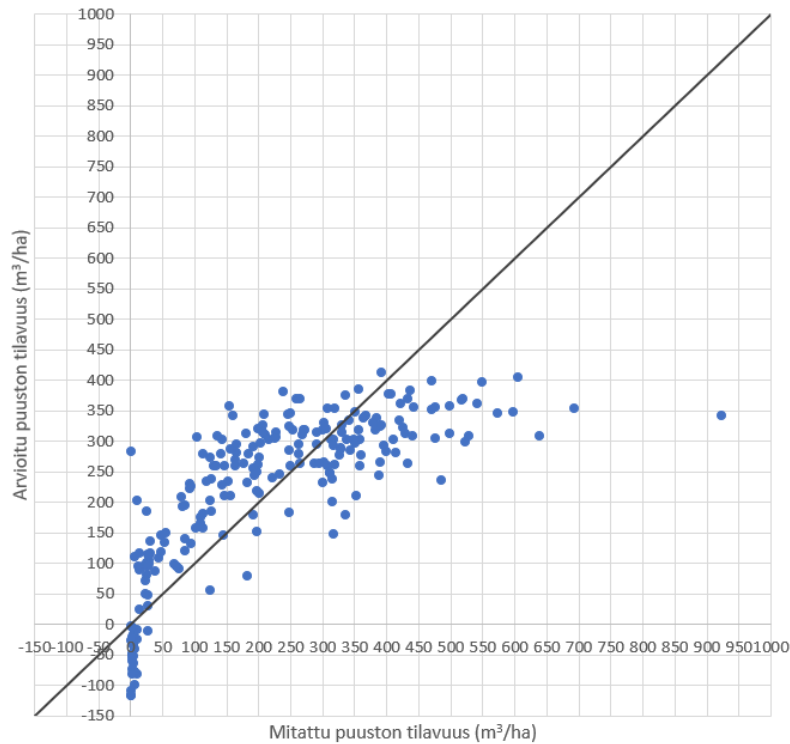


Kuva 3. Puulajiluokan männyn regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,45, RMSE 68,5  $m^3/ha$ , nRMSE 45,1 %, MdAE 44,4  $m^3/ha$ , nMdAE 30,3 %, maksimitilavuus 278,5  $m^3/ha$  sekä minimitilavuus -77,5  $m^3/ha$ .

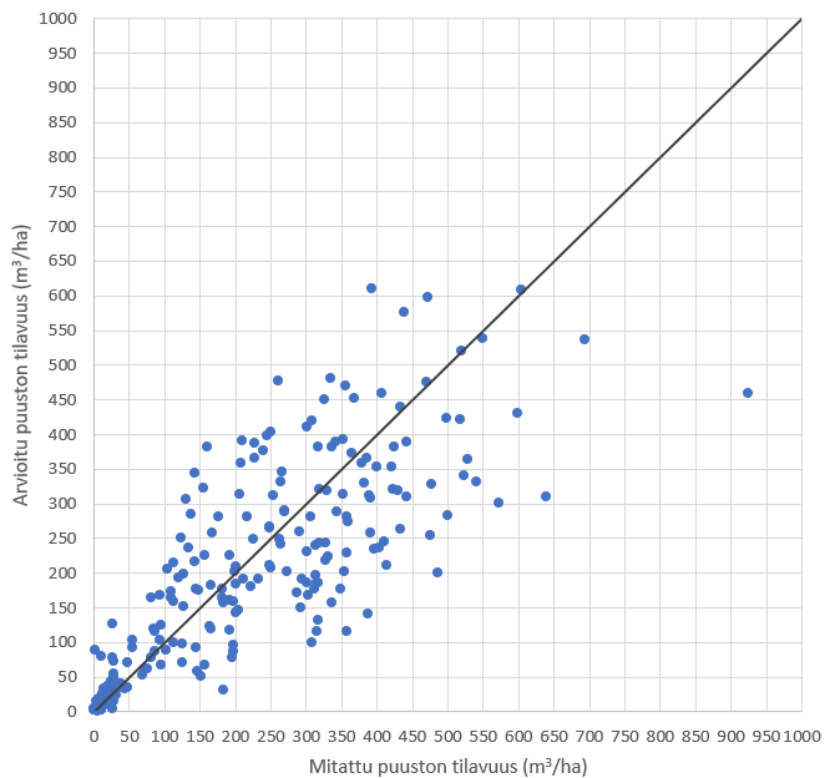


Kuva 4. Puulajiluokan männyn regressiomalli laskettu kanavalla B5, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,55, RMSE 75,6  $m^3/ha$ , nRMSE 49,8 %, MdAE 37,7  $m^3/ha$ , nMdAE 25,7 %, maksimitilavuus 471,9  $m^3/ha$  sekä minimitilavuus 5,5  $m^3/ha$ .

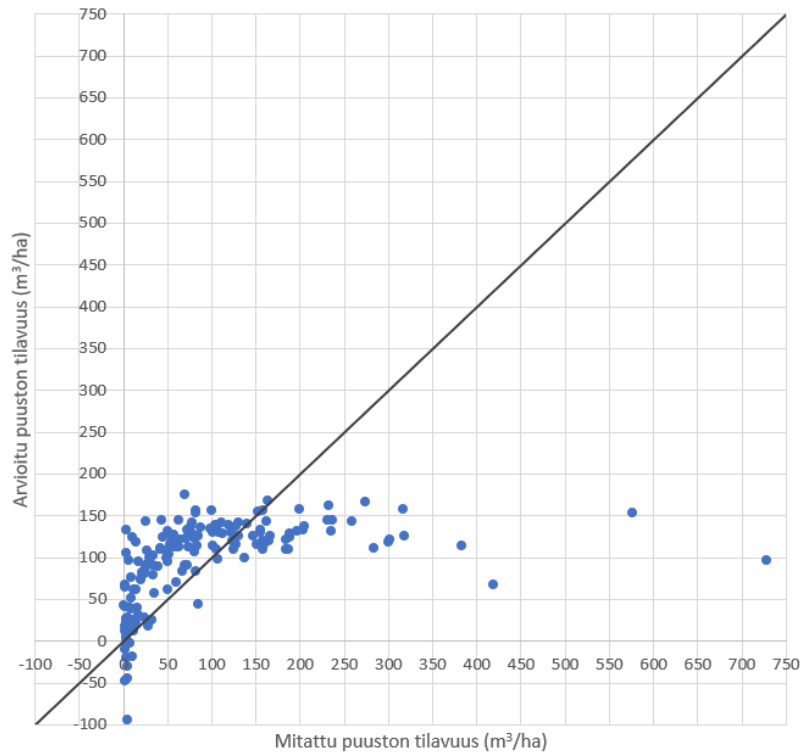




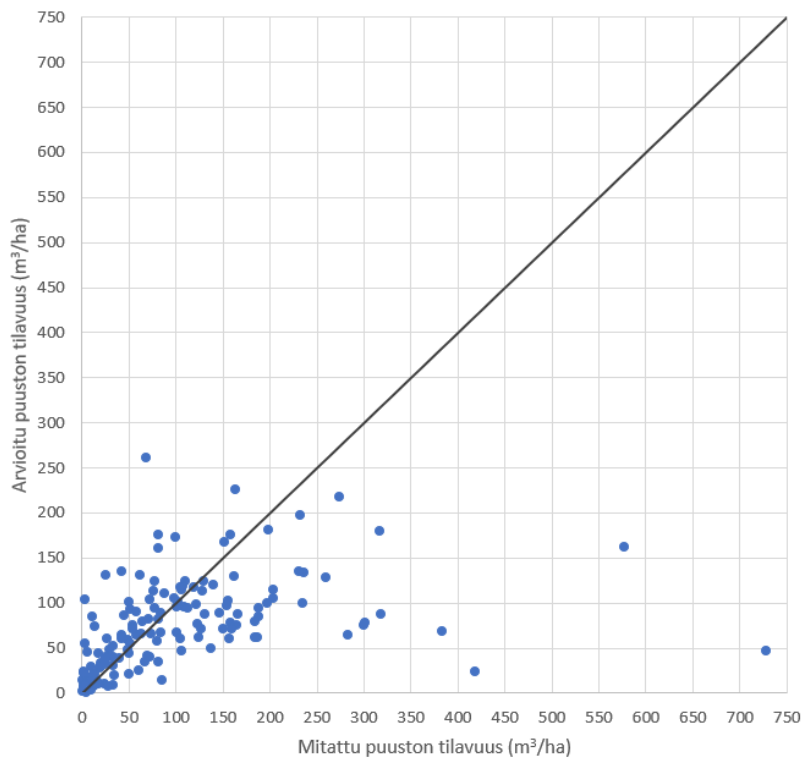
Kuva 5. Puulajiluokan kuusen regressiomalli laskettu kanavilla B3 ja B7, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,59, RMSE 108,6 m<sup>3</sup>/ha, nRMSE 49,4 %, MdAE 70,5 m<sup>3</sup>/ha, nMdAE 35,2 %, maksimitilavuus 413,0 m<sup>3</sup>/ha sekä minimitilavuus -117,7 m<sup>3</sup>/ha.



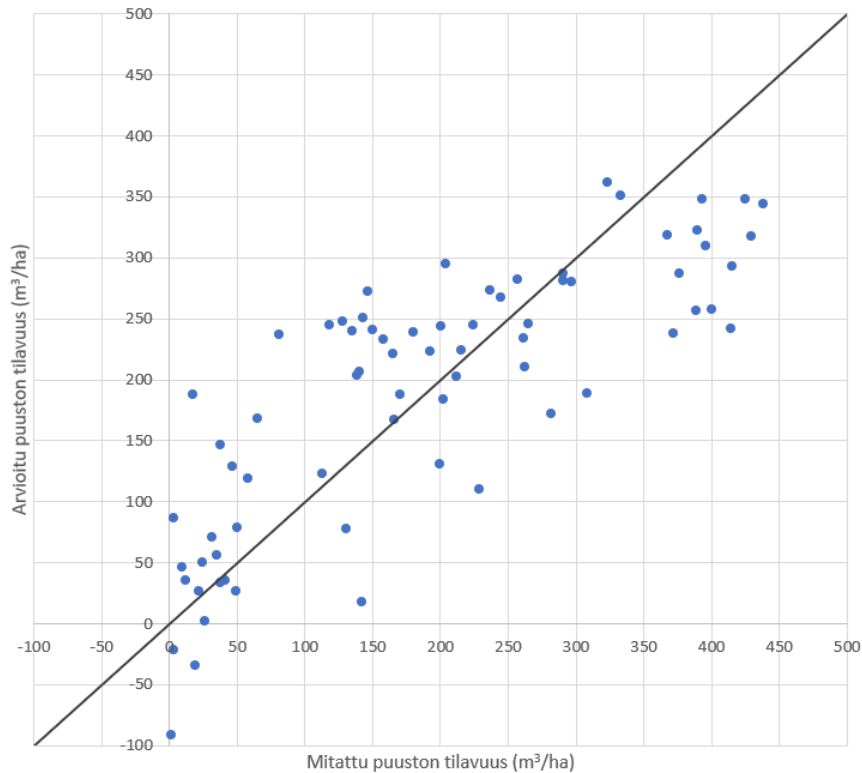
Kuva 6. Puulajiluokan kuusen regressiomalli laskettu kanavalla B5, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,84, RMSE 103,8 m<sup>3</sup>/ha, nRMSE 47,2 %, MdAE 52,2 m<sup>3</sup>/ha, nMdAE 26,0 %, maksimitilavuus 610,7 m<sup>3</sup>/ha sekä minimitilavuus 1,4 m<sup>3</sup>/ha.



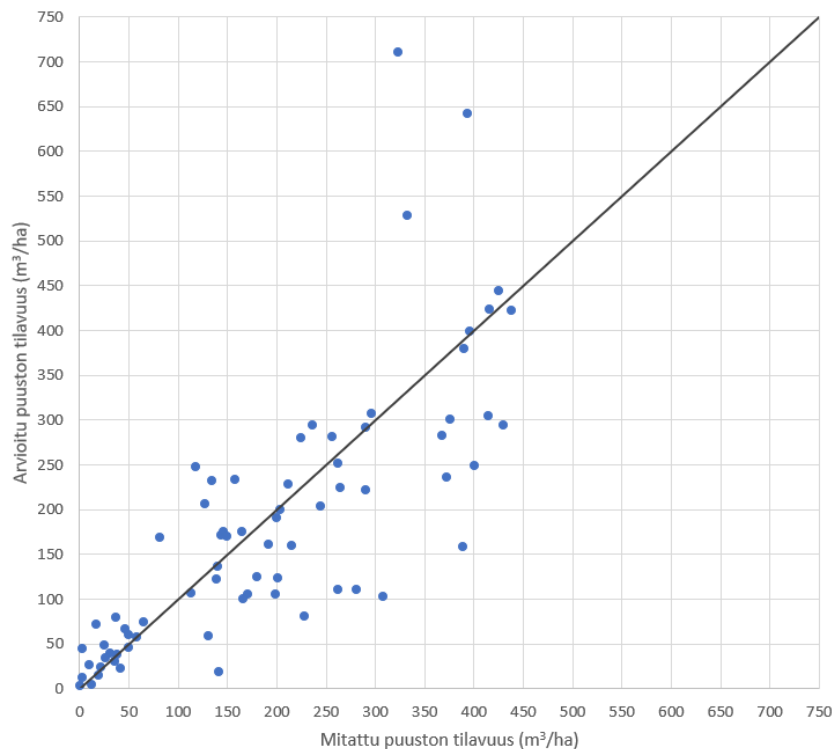
Kuva 7. Puulajiluokan lehtipuun regressiomalli laskettu kanavalla B5, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,23, RMSE 93,0  $\text{m}^3/\text{ha}$ , nRMSE 98,3 %, MdAE 44,6  $\text{m}^3/\text{ha}$ , nMdAE 65,6 %, maksimitilavuus 174,6  $\text{m}^3/\text{ha}$  sekä minimitilavuus -95,0  $\text{m}^3/\text{ha}$ .



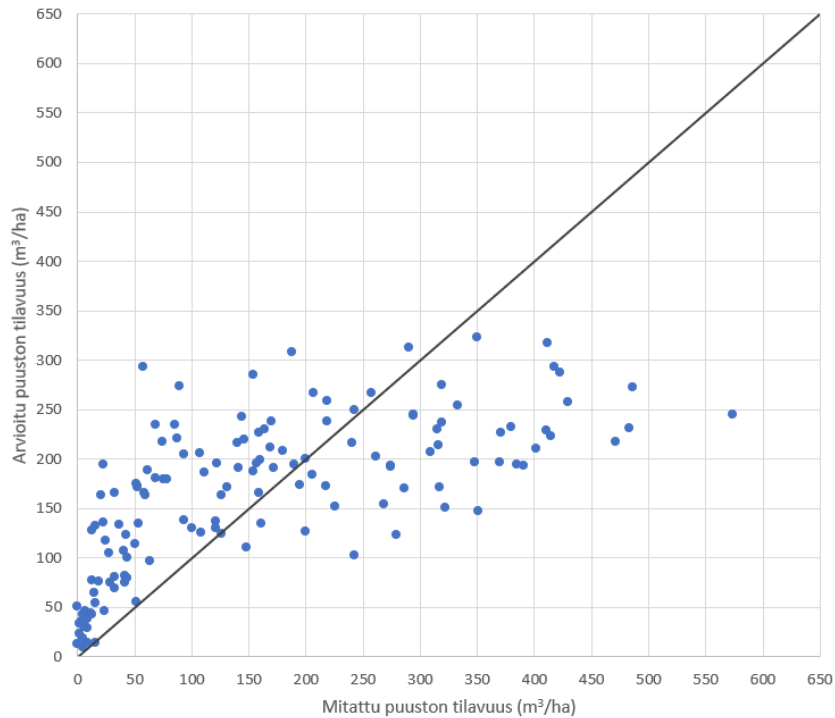
Kuva 8. Puulajiluokan lehtipuun regressiomalli laskettu kanavalla B5, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,59, RMSE 97,7  $\text{m}^3/\text{ha}$ , nRMSE 103,2 %, MdAE 20,3  $\text{m}^3/\text{ha}$ , nMdAE 29,8 %, maksimitilavuus 260,5  $\text{m}^3/\text{ha}$  sekä minimitilavuus 0,7  $\text{m}^3/\text{ha}$ .



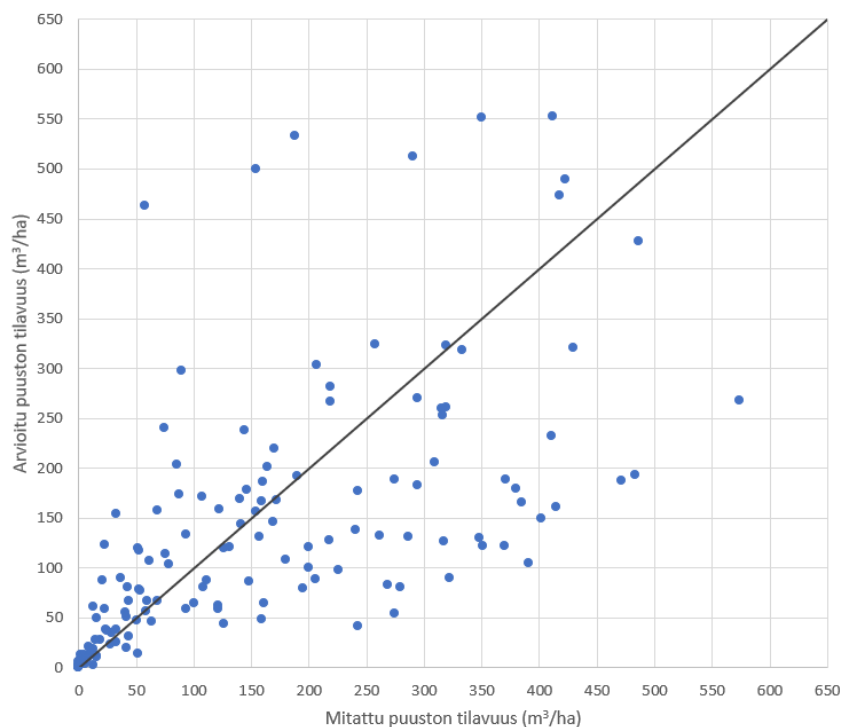
Kuva 9. Puulajiluokan havusekametsän regressiomalli laskettu kanavalla B7, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,64, RMSE 78,8 m<sup>3</sup>/ha, nRMSE 41,3 %, MdAE 54,9 m<sup>3</sup>/ha, nMdAE 31,4 %, maksimitilavuus 362,0 m<sup>3</sup>/ha sekä minimitilavuus -91,5 m<sup>3</sup>/ha.



Kuva 10. Puulajiluokan havusekametsän regressiomalli laskettu kanavilla B3 ja B6, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,73, RMSE 93,5 m<sup>3</sup>/ha, nRMSE 49,0 %, MdAE 30,2 m<sup>3</sup>/ha, nMdAE 17,2 %, maksimitilavuus 710,0 m<sup>3</sup>/ha sekä minimitilavuus 3,0 m<sup>3</sup>/ha.



Kuva 11. Puulajiluokan sekametsän regressiomalli laskettu kanavilla B5 ja B6, kun selitettävänä muuttujana oli tilavuus. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,47, RMSE 101,5 m<sup>3</sup>/ha, nRMSE 68,5 %, MdAE 65,9 m<sup>3</sup>/ha, nMdAE 60,3 %, maksimitilavuus 323,4 m<sup>3</sup>/ha sekä minimitilavuus -157,7 m<sup>3</sup>/ha.



Kuva 12. Puulajiluokan sekametsän regressiomalli laskettu kanavalla B5, kun selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi. Mallin tunnusluvut olivat korjattu  $R^2$  -luku 0,72, RMSE 115,0 m<sup>3</sup>/ha, nRMSE 77,6 %, MdAE 38,1 m<sup>3</sup>/ha, nMdAE 34,8 %, maksimitilavuus 552,9 m<sup>3</sup>/ha sekä minimitilavuus 0,4 m<sup>3</sup>/ha.

## Liite 4. Regressioanalyysin validoinnin tulokset

*Taulukko 1. Koko aineiston validoinnin tulokset, kun regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävinä muuttujina kanavat B5 ja B6.*

Opetus- aineisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,002488*B5-0,000137*B6+3,448201)}$	0,72	102,2	63,0	38,2	28,0	807,1	0,7
2	$10^{(-0,002373*B5-0,000169*B6+3,438848)}$	0,73	101,9	62,5	37,6	27,3	750,9	1,5
3	$10^{(-0,002613*B5-0,000105*B6+3,461134)}$	0,70	103,9	64,2	36,0	26,4	835,3	0,6

*Taulukko 2. Puulajiluokan männyn validoinnin tulokset, kun regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävinä muuttujana kanava B5.*

Opetus- aineisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,002608*B5+3,323264)}$	0,55	75,6	49,8	37,7	25,7	471,9	5,5
2	$10^{(-0,002603*B5+3,319527)}$	0,56	72,2	45,0	38,2	24,1	417,5	7,4
3	$10^{(-0,002613*B5+3,327237)}$	0,53	78,7	54,9	36,8	27,1	474,9	5,5

*Taulukko 3. Puulajiluokan kuusen validoinnin tulokset, kun regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävinä muuttujana kanava B5.*

Opetus- aineisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,002778*B5+3,275415)}$	0,84	103,8	47,2	52,2	26,0	610,7	1,4
2	$10^{(-0,002851*B5+3,304553)}$	0,83	114,3	54,1	56,6	28,5	634,0	1,8
3	$10^{(-0,002705*B5+3,245833)}$	0,85	91,2	39,7	54,2	25,1	585,8	1,6

*Taulukko 4. Puulajiluokan lehtipuun validoinnin tulokset, kun regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävinä muuttujana kanava B5.*

Opetusai- neisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,002741*B5+3,317263)}$	0,59	97,7	103,2	20,3	29,8	260,5	0,7
2	$10^{(-0,002674*B5+3,298341)}$	0,56	67,1	77,6	21,2	31,6	183,4	0,8
3	$10^{(-0,002802*B5+3,332513)}$	0,61	120,4	117,5	19,0	27,5	257,7	2,5

*Taulukko 5. Puulajiluokan havusekametsän validoinnin tulokset, kun regressiomallin selitettävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävinä muuttujina kanavat B3 ja B6.*

Opetus- aineisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,001666*B3-0,000819*B6+3,811811)}$	0,73	93,5	49,0	30,2	17,2	710,0	3,0
2	$10^{(-0,002348*B3-0,000582*B6+3,588914)}$	0,74	80,5	40,5	34,9	18,3	395,2	15,1
3	$10^{(-0,001654*B3-0,001023*B6+4,116982)}$	0,73	149,8	81,5	55,8	33,2	907,8	1,5

*Taulukko 6. Puulajiluokan sekametsän validoinnin tulokset, kun regressiomallin selittävänä muuttujana oli tilavuuden luonnollinen logaritmi ja selittävänä muuttujana kanava B5.*

Opetusai- neisto	Malli	Kor- jattu R <sup>2</sup>	RMSE (m <sup>3</sup> /ha)	nRMSE (%)	MdAE (m <sup>3</sup> /ha)	nMdAE (%)	Max V (m <sup>3</sup> /ha)	Min V (m <sup>3</sup> /ha)
1	$10^{(-0,003319*B5+3,596441)}$	0,72	115,0	77,6	38,1	34,8	552,9	0,4
2	$10^{(-0,003598*B5+3,739435)}$	0,74	125,5	92,3	39,8	44,0	651,4	0,3
3	$10^{(-0,003166*B5+3,512060)}$	0,70	111,1	66,7	32,8	26,0	481,5	0,5